

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE INFORMÁTICA

Departamento de Ingeniería del Software e Inteligencia Artificial



**COMBINACIÓN DE CLASIFICADORES PARA
IDENTIFICACIÓN DE TEXTURAS EN IMÁGENES
NATURALES: NUEVAS ESTRATEGIAS LOCALES
Y GLOBALES.**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR**

María Guijarro Mata-García

Bajo la dirección de los doctores

Gonzalo Pajares Martinsanz
Luis Garmendia Salvador

Madrid, 2010

- ISBN: 978-84-693-1122-6

UNIVERSIDAD COMPLUTENSE DE MADRID
Facultad de Informática



Departamento de Ingeniería del Software e Inteligencia Artificial

TESIS DOCTORAL

COMBINACIÓN DE CLASIFICADORES PARA IDENTIFICACIÓN DE TEXTURAS EN IMÁGENES NATURALES: nuevas estrategias locales y globales

Memoria para optar al grado de doctor presentada por:

Dña. María Guijarro Mata-García

Directores:

Dr. D. Gonzalo Pajares Martinsanz

Dr. D. Luis Garmendia Salvador

Madrid (España), 2009

Agradecimientos

Hace tres años decidí continuar estudiando e intentar alcanzar el sueño que supone esta tesis. Fue entonces cuando coincidí con Gonzalo Pajares. Sería injusto agradecerle tan sólo este trabajo, ya que todo lo conseguido desde entonces se lo debo. Gracias a tu ayuda, tu paciencia, tu comprensión y sobre todo gracias a la confianza que has depositado en mí. Nunca te has cansado a la hora de animarme, siempre me has escuchado y me has enseñado, no sólo a mejorar profesionalmente sino también como persona. Espero me dejes seguir aprendiendo a tu lado.

También agradecer a Luis Garmendia su codirección.

En Dimap encontré ese apoyo que siempre se necesita y un impulso para continuar. Ángel y Felipe, gracias también por las imágenes que me habéis dejado para poder realizar este trabajo.

Durante estos dos años en el CES Felipe II he tenido la oportunidad de desarrollarme profesionalmente tanto en mi docencia como en mi investigación, para lo cual he contado con los consejos y el apoyo de mis compañeros, gracias Chema, Nuria y Fernando; y gracias Iñaki por confiar en mi capacidad y recordarme que éste es el camino a seguir con mucho esfuerzo.

Ángela, muchas gracias por haberme dado la oportunidad de participar en esos proyectos punteros, que tan brillantemente diriges en el IAI.

Un profundo agradecimiento al grupo ISCAR con Jesús, Pepe y de nuevo Gonzalo, por el apoyo y confianza otorgados.

Sin la comprensión primero de Marta y después de José Ramón, que me han prestado parte de su espacio de trabajo, me hubiese resultado difícil llegar hasta aquí.

Carmen, María, Rubén y Guille, gracias por escucharme, animarme y darme siempre el punto de vista positivo, que muchas veces este camino sabe esconder tan bien.

Nada de todo esto se habría hecho posible sin la educación y la enseñanza que con tanto esmero he recibido de mis padres, mis dos grandes referencias en la vida. Gracias Chus,

gracias por tantas noches sin dormir a mi lado, gracias por tanto consuelo, apoyo y cariño, y por haber sido además de una madre ejemplar la mejor amiga que una hija podría esperar. Y Nano, gracias por todas las tardes que me has dedicado, por enseñarme a valorar el esfuerzo, enseñarme a querer el estudio; gracias por haber sido mi confidente desde que tengo uso de razón y haber luchado siempre a mi lado por conseguir llegar a ser tanto una buena profesional como una buena persona. Gracias de nuevo a los dos por haberme dado la oportunidad de llegar a este día.

María y Antonio gracias por todo vuestro amor.

David, nunca te has rendido, siempre has estado conmigo. En los momentos más duros has dejado todo para estar a mi lado, para cuidarme. Siempre me has animado para seguir adelante, para seguir luchando, convenciéndome de que con trabajo y esfuerzo podríamos con todo. Pues parece que hemos podido con un poquito más, gran parte de este trabajo te lo debo.

Por último, ¿Qué te voy a decir a ti? Que me conoces mejor que nadie, que has sabido lo importante que es para mi este trabajo y has sabido ayudarme y apoyarme siempre. Gracias por tantas noches de escucha y por tanto amor. Gracias por prestarme tu estrella. Las dos sabemos que nada va a cambiar.

Madrid, Marzo 2009

Resumen

La combinación de clasificadores constituye hoy en día un campo emergente en los sistemas de toma de decisiones. Estos sistemas se caracterizan por su capacidad para el tratamiento de grandes cantidades de información, dentro de lo que se denomina minería de datos. Un área de interés dentro de este ámbito es la identificación de texturas en imágenes, particularmente texturas naturales de exterior. Es aquí donde se centra el trabajo de investigación que se presenta.

La motivación de la investigación está suscitada por un interés empresarial creciente para la identificación de las texturas en imágenes procedentes de sensores aerotransportados, cuya finalidad última está encaminada hacia aplicaciones concretas, entre las que destacan, la ordenación del territorio o la prevención de posibles catástrofes, tales como incendios forestales. La gran cantidad de información proporcionada por esos sensores debe ser convenientemente tratada.

El objetivo principal consiste en el diseño de estrategias de clasificación capaces de mejorar los resultados proporcionados por las herramientas comerciales existentes y en la medida de lo posible automatizar el proceso de clasificación.

Como estrategia de clasificación se propone la combinación de clasificadores, siguiendo la tendencia actual de las investigaciones en este campo. Bajo la perspectiva de las imágenes se proponen dos estrategias combinadas a nivel de píxel, una de naturaleza local y otra global. La primera aborda el problema de forma específica, clasificando cada píxel en la imagen de forma individual teniendo en cuenta sus propiedades espectrales, mientras que la segunda considera esas mismas propiedades del píxel a clasificar junto con las de sus vecinos, así como las clasificaciones previas de los píxeles en la vecindad.

Bajo el enfoque local se proponen sendos procedimientos de combinación bajo la perspectiva de la lógica fuzzy, el primero de ellos se basa en el paradigma de la *teoría de la decisión multi-criterio* y el segundo en la *integral fuzzy*.

Dentro del enfoque global, se proponen a su vez dos procedimientos de combinación, el primero basado en el concepto de los *mapas cognitivos fuzzy* y el segundo bajo el paradigma de la *relajación probabilista*.

Para llevar a cabo la combinación se utilizan seis clasificadores clásicos en su versión original, como son: el método de agrupamiento borroso, el estimador paramétrico de máxima verosimilitud, el estimador no paramétrico de la ventana de Parzen, los mapas auto-organizativos y dos versiones derivadas del método de cuantización vectorial.

La comparación de los resultados frente a estos clasificadores individuales, así como frente a otros métodos clásicos combinados, permite verificar que las estrategias de combinación propuestas en este trabajo mejoran los resultados, quedando de este modo justificadas.

Adicionalmente, se ha diseñado una estrategia de naturaleza no supervisada a partir de un método supervisado. Esto permite el tratamiento de los datos de forma automática, en concordancia con el objetivo inicial.

Todos los métodos se han diseñado con una filosofía de diseño abierta, en el sentido de que permiten la incorporación de nuevos clasificadores individuales o la variación de los mecanismos de combinación, particularmente en los métodos globales. Con este planteamiento, se pretende la consecución de un doble objetivo en el futuro, a saber: la mejora de los resultados en la clasificación de las texturas en imágenes, lo que todavía puede ser posible y la utilización de los métodos combinados en otro tipo de aplicaciones, incluso fuera del ámbito de las imágenes.

Palabras clave: clasificación no supervisada, texturas naturales, combinación de clasificadores, toma de decisiones multi-criterio, integral fuzzy, mapas cognitivos fuzzy, relajación probabilista.

Abstract

Nowadays the classifier combination is an emerging area in the decision making systems. These systems are designed for dealing with large-scale information systems, under data mining approaches. The identification of textures in images is a field of interest, particularly in outdoor natural images. We have focused the interest in this field.

The motivational research is aroused by a business growing interest for texture identification in images supplied by airborne sensors; where, among others, specific applications are: land use planning or possible catastrophic fire forest prevention. The great quantity of information provided by these sensors must be conveniently processed.

The main goal of this work is the design of classification strategies, which are able to improve the results provided by the existing commercial tools and as far as possible, to automate the classification process.

The classification strategy proposed is the combination of classifiers, according to the tendency of the research in this field. From the point of view of the images, two strategies are proposed at the pixel level, namely: local and global. Under the first, each pixel in the image is classified individually based on its spectral properties. In the second, which is iterative, not only the spectral properties of each pixel to be classified are considered, but also the spectral properties of each neighbour pixel and besides the classification results of the neighbours during the previous iteration.

The local approach includes two combination methods under the fuzzy logic, the first is based on the *multi-criteria decision making* paradigm and the second on the *fuzzy integral*.

Under the global approach, also two combined processes are proposed; the first applies the *fuzzy cognitive maps* and the second the *probabilistic relaxation* paradigm.

The combinations are based on six classical classifiers under their original version, they are: the fuzzy clustering, the parametric maximum likelihood estimator, the non-parametric Parzen's window estimator, the self-organizing maps and two versions of the learning vector quantization.

The results are compared favourably against these individual classifiers, and also against other classical combined methods. This allows us to verify that the proposed approaches outperform the results of classical combiners and the individual classifiers. This justifies its choice.

Additionally, an unsupervised strategy has been designed from a supervised method. This makes that the data processing becomes automatic in accordance to the initial objective.

All methods have been proposed under an open design structure, allowing the addition of new individual classifiers or even though the change of the combination schemes, particularly for the global methods. Based on this, we try to achieve a twofold objective in the future, namely: the improving of the results, because they could be probably improved and the application of the proposed strategies for solving different problems out of the scope of the images.

Keywords: unsupervised classification, natural textures, combining classifiers, multi-criteria decision making, fuzzy integral, fuzzy cognitive maps, probabilistic relaxation.

Índice general

Capítulo 1. Introducción	1
1.1 Antecedentes	1
1.2 Identificación del problema	7
1.3 Motivación y Objetivos	9
1.3.1 Motivación	9
1.3.2 Objetivos	12
1.4 Metodología	12
1.5 Aportaciones de la investigación	13
1.6 Organización de la tesis	15
 Capítulo 2. Estado del arte	 17
2.1 Visión general	17
2.2 Métodos de clasificación de texturas en imágenes	18
2.3 Métodos de clasificación combinados	23
2.3.1 Clasificadores combinados de índole general	23
2.3.2 Combinación de clasificadores en texturas de imágenes	24
2.4 Características de las imágenes en la clasificación de texturas naturales	27
2.4.1 Características espectrales a nivel de píxel	28
2.4.2 Características con patrones espaciales	31
2.4.3 Combinación de características espectrales y espaciales	33
2.5 Métodos de clasificación mediante información espacial	35
2.5.1 Información espacial mediante Campos Aleatorios de Markov	36
2.5.2 Relaciones estructurales	39
2.5.3 Otros enfoques	40
2.6 Resumen del análisis bibliográfico	41

Capítulo 3. Clasificadores clásicos: individuales y combinados	45
3.1 Introducción	45
3.2 Fases de entrenamiento y decisión	46
3.2.1 Procesos comunes	47
3.2.2 Procesos diferenciados	48
3.3 Métodos de clasificación clásicos individuales	49
3.3.1 Agrupamiento borroso	49
3.3.1.1 Fase de entrenamiento	49
3.3.1.2 Fase de decisión	54
3.3.2 Estimación Paramétrica: máxima verosimilitud	54
3.3.2.1 Fase de entrenamiento	54
3.3.2.2 Fase de decisión	58
3.3.3 Estimación no paramétrica: ventana de Parzen	59
3.3.3.1 Fase de entrenamiento	59
3.3.3.2 Fase de decisión	63
3.3.4 Mapas Auto-Organizativos	64
3.3.4.1 Fase de entrenamiento	64
3.3.4.2 Fase de decisión	67
3.3.5 Cuantización vectorial	68
3.3.5.1 Fase de entrenamiento	68
3.3.5.2 Fase de decisión	70
3.3.6 Resumen de clasificadores individuales	70
3.4 Funciones criterio para la partición	72
3.4.1 Coeficiente de Partición	73
3.4.2 Suma del error al cuadrado	74
3.4.3 Criterio de varianza mínima relacionada	76
3.4.4 Comportamiento de las funciones criterio	77
3.4.5 Obtención de las competencias de los clasificadores individuales	80
3.5 Métodos de clasificación combinados	81
3.5.1 Terminología y taxonomías	82
3.5.2 Métodos para combinar las salidas de los clasificadores	84
3.5.2.1 Votación Mayoritaria	85
3.5.2.2 Combinación de Bayes	86
3.5.2.3 Combinación mediante funciones	87

3.5.2.4	Combinación mediante operadores de agregación fuzzy	90
3.5.2.5	Resumen de los modos de combinación de clasificadores	91
3.5.2.6	Combinación mediante la integral fuzzy	93
3.6	Resumen del capítulo	96
Capítulo 4. Estrategias locales y globales de combinación de clasificadores		97
4.1	Introducción	97
4.2	Diseño de un clasificador local no supervisado: MCDMF	100
4.2.1	Fase de entrenamiento	101
4.2.2	Descripción general del paradigma MCDM fuzzy	103
4.2.2.1	Formulación del problema y definiciones previas	103
4.2.2.2	Normalización	105
4.2.2.3	Proceso de decisión	106
4.2.3	Cálculo de las competencias de los clasificadores individuales	108
4.2.4	Fase de decisión: combinación de clasificadores	110
4.3	Diseño de un clasificador local no supervisado: Integral Fuzzy	114
4.4	Diseño de un clasificador global no supervisado: Mapas Cognitivos Fuzzy	116
4.4.1	Fase de entrenamiento	117
4.4.2	Mapas Cognitivos Fuzzy: descripción general	117
4.4.3	Fase de decisión	120
4.4.3.1	Topología de los MCFs	121
4.4.3.2	Criterio de decisión	122
4.4.3.3	Proceso de actualización	122
4.4.3.4	Definición de los pesos de conexión	125
4.4.3.5	Definición del factor de decadencia	130
4.4.3.6	Síntesis del proceso global basado en los MCFs	131
4.5	Combinación de clasificadores por relajación	133
4.6	Resumen del capítulo	136
Capítulo 5. Análisis de resultados		139
5.1	Introducción	139
5.2	Descripción de los datos utilizados	141

5.2.1	Descripción de los datos genéricos	141
5.2.2	Descripción de las imágenes aéreas	144
5.3	Ajuste de parámetros	148
5.3.1	Fase de entrenamiento	149
5.3.1.1	Parámetros del clasificador borroso	150
5.3.1.2	Parámetros del estimador de la ventana de Parzen	151
5.3.1.3	Mapas auto-organizativos	153
5.3.1.4	Cuantización vectorial	154
5.3.1.5	Umbral de las particiones	155
5.3.2	Fase de decisión	156
5.3.2.1	Mapas Cognitivos Fuzzy y relajación	156
5.3.2.2	Operador de agregación de Yager	157
5.3.3	Resumen de parámetros	157
5.4	Diseño de una estrategia general de pruebas	158
5.5	Entrenamiento inicial	162
5.5.1	Estimación inicial del número de clases y competencias de los clasificadores	162
5.5.2	Aprendizaje de los parámetros de cada clasificador	167
5.6	Generación de las clases de referencia	169
5.7	Resultados obtenidos	172
5.7.1	Cómputo del porcentaje de error	173
5.7.2	Comportamiento general de los diferentes métodos	174
5.7.3	Análisis de los tiempos de cómputo	179
5.7.3.1	Fase de decisión: métodos globales	179
5.7.3.2	Fase de decisión: métodos locales combinados e individuales	180
5.7.3.3	Fase de entrenamiento: métodos individuales	181
5.7.4	Análisis comparativo de resultados	182
5.7.4.1	Influencia del aprendizaje	182
5.7.4.2	Clasificadores individuales	183
5.7.4.3	Clasificadores combinados locales clásicos	184
5.7.4.4	Clasificadores combinados globales	185
5.7.4.5	Clasificadores combinados locales basados en lógica fuzzy	185
5.7.4.6	Análisis global	186
5.7.5	Ilustración de resultados relevantes	187
5.8	Resumen del capítulo	191

Capítulo 6. Conclusiones y trabajo futuro	193
6.1 Introducción	193
6.2 Conclusiones	193
6.2.1 Estrategias combinadas propuestas	193
6.2.2 Avances en la combinación de clasificadores	194
6.3 Trabajo futuro	195
6.3.1 Métodos combinados locales	195
6.3.2 Métodos combinados globales	196
6.3.3 Ajuste de parámetros	197
6.3.4 General	197
Bibliografía	199

Índice de Abreviaturas

AVIRIS	Airborne Visible InfraRed Imaging Spectrometer
BC	Base de Conocimiento
BN	estimador no paramétrico basado en la ventana de Parzen
BP	estimador paramétrico de máxima verosimilitud
CEDEX	Centro de Estudios y Experimentación de Obras Públicas
CICYT	Centro de Investigación Científica y Tecnológica
CIFOR	Centro de Investigación Forestal
CSIC	Consejo Superior de Investigaciones Científicas
EM	Expectation Maximization
FA	Agregación fuzzy de Yager
Fps	frames per second
HSI	H, hue, S, saturation e I, Intensity
IF	Integral Fuzzy
INIA	Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria
INTA	Instituto Nacional de Técnica Aeroespacial
ISCAR	Ingeniería de Sistemas, Control, Automatización y Robótica
k-NN	k-Nearest Neighbours
L1 y L2	métodos de cuantización vectorial “Learning Vector Quantization”
LVQ	Learning Vector Quantization
MA	Regla del Máximo
MCDM	Multi-Criteria Decision Making
MCDMF, MC	Multi-Criteria Decision Making Fuzzy
MCF, MF	Mapas Cognitivos Fuzzy
MDT	Mapas Digitales del Terreno
ME	Regla de la Media aritmética
MI	Regla del Mínimo
MRFs	Markov Random Fields
MV	Votador Mayoritario
PC	Coefficiente de Partición
PR	Regla del Producto

REL, RL	esquemas de relajación
RGB	R, red; G, green y B, blue
SAR	Synthetic Aperture Radar
SE	suma del error al cuadrado
SIC	Satellite Image Corporation
SITGA	Servicio Territorial de Galicia
SOM, SO	Mapas Auto-Organizativos o Self Organizing feature Maps
TOPSIS	Technique for Order Performance by Similarity to Ideal Solution
UTM	Universal Transverse Mercator
UAVs	Unmanned Aerial Vehicles
VC	criterio varianza mínima relacionada

Índice de Símbolos

\mathbf{x}_k	valor del píxel k en el modelo de color RGB
w_i	clase i
μ_{ik}	grado de pertenencia de \mathbf{x}_k en la clase w_i
\mathbf{v}_j	centro de la clase w_j
m	peso exponencial
t_{\max}	número máximo de iteraciones
ε	criterio de convergencia
\mathbf{m}_j	media de la clase w_j
$p(\mathbf{x} \mathbf{m}_j, C_j)$	probabilidad de pertenencia de \mathbf{x} a la clase w_j
$P(\mathbf{x} / \mathbf{m}_j, C_j)$	probabilidad total
$K(\cdot)$	ventana o función del núcleo
r	coeficiente del ancho de banda
h	parámetro de suavizado o ancho de la ventana
$d(\mathbf{x}, \mathbf{x}_j)$	distancia métrica.
c_{ij}	peso de conexión entre nodos
$\gamma(t=0)$	razón de aprendizaje
α_{inicial}	ancho inicial de la vecindad
α_{final}	ancho final de la vecindad
$PD(\mathbf{x})$	perfil de decisión
c	número de clases
A_j	alternativa j
CR_i	criterio i
Z	vector de pesos específicos asociados a cada criterio
$\hat{\Delta}^{SE}(\hat{c}), \hat{\Delta}^{VC}(\hat{c})$	variación relativa de SE y VC
$\alpha_l^{SE}(\hat{c}_i), \alpha_{BN}^{VC}(\hat{c}_i)$	grados de competencia absoluta

Índice de Figuras

Figura 1.1	(a) Imagen aérea original; (b) identificación de las texturas; (c) camino óptimo	5
Figura 1.2	Diversidad de texturas naturales	8
Figura 2.1	Diferentes ejemplos de texturas procedentes de las bases de imágenes Brodatz y VisTex	31
Figura 2.2	Ejemplo de imagen aérea con diferentes texturas naturales	33
Figura 2.3	Relación de vecindad en un Campo Aleatorio de Markov	36
Figura 3.1	Esquema general de un proceso aprendizaje-decisión	47
Figura 3.2	Variación de h_j frente a r para valores dados de h y n_j	62
Figura 3.3	Modelo de arquitectura de Kohonen en el caso bi-dimensional	65
Figura 3.4	Dos agrupamientos con diferentes valores del coeficiente SE	75
Figura 3.5	(a) Imagen original, (b) distribución de las muestras en cuatro clases.	76
Figura 3.6	Variación relativa de los coeficientes PC , SE y VC entre números de clases consecutivos	79
Figura 3.7	Niveles en la combinación de clasificadores	83
Figura 3.8	Combinación de clasificadores individuales	92
Figura 3.9	Esquema de funcionamiento de la integral fuzzy en la combinación de clasificadores	94
Figura 4.1	Diseño de un clasificador no supervisado combinado: fases de entrenamiento y decisión	101
Figura 4.2	Arquitectura de un clasificador híbrido basado en el enfoque MCF: entrenamiento y decisión	116
Figura 4.3	Un ejemplo sencillo de Mapa Cognitivo fuzzy	118
Figura 4.4	Esquema gráfico sobre el método de combinación basado en el esquema MCF para cuatro clases	133
Figura 5.1	Imágenes hemisféricas de zonas boscosas (a) tomada en un día soleado; (b) tomada en un día nublado	142
Figura 5.2	Imágenes representativas de los conjuntos utilizados: (a) imagen perteneciente al grupo de entrenamiento inicial SI; (b) imagen representativa del conjunto testigo S0	147
Figura 5.3	Imágenes representativas de los conjuntos utilizados: (a) S1; (b) S2 y (c) S3	148
Figura 5.4	Síntesis de la estrategia de prueba	162

Figura 5.5	Valores de $\hat{\Delta}_{BP}^{SE}(c)$ y $\hat{\Delta}_{BN}^{VC}(c)$ frente al número de clases	164
Figura 5.6	(a) imagen de textura natural correspondiente al conjunto S_1 ; (b) distribución tri-dimensional de las muestras con sus centros de clase correspondientes	168
Figura 5.7	Distribución bi-dimensional de los mismos datos representados en la figura 5.6 (b)	169
Figura 5.8	(a) imagen original perteneciente al conjunto S_0 ; (b) correspondencia entre etiquetas y clases; (c) imagen etiquetada en sus cuatro clases de acuerdo a la equivalencia establecida en (b)	170
Figura 5.9	Clase 1: (a) imagen binaria obtenida mediante el clasificador BP; (b) imagen de referencia	171
Figura 5.10	Clase 2: (a) imagen binaria obtenida mediante el clasificador BP; (b) imagen de referencia	171
Figura 5.11	Clase 3: (a) imagen binaria obtenida mediante el clasificador BP; (b) imagen de referencia	172
Figura 5.12	Clase 4: (a) imagen binaria obtenida mediante el clasificador BP; (b) imagen de referencia	172
Figura 5.13	Porcentaje medio de error de los clasificadores individuales para S_0 en los tres PASOS	177
Figura 5.14	Porcentaje medio de error de los clasificadores combinados clásicos para S_0 en los tres PASOS	177
Figura 5.15	Porcentaje medio de error de los clasificadores combinados iterativos para el conjunto S_0 en los tres PASOS	178
Figura 5.16	Porcentaje medio de error de los clasificadores combinados basados en lógica fuzzy para el conjunto S_0 en los tres PASOS	178
Figura 5.17	Porcentaje medio de error de los clasificadores que han proporcionado el mejor resultado dentro de su categoría para el conjunto S_0 en los tres PASOS	179
Figura 5.18	Porcentaje medio de error de los clasificadores que han proporcionado el mejor resultado dentro de su categoría para los conjuntos S_1 , S_2 y S_3 en los tres PASOS	179
Figura 5.19	Imagen original perteneciente al conjunto S_3	187
Figura 5.20	Resultados de la clasificación para la imagen de la figura 5.19 obtenidos mediante: (a) el clasificador FC y (b) el clasificador BP	188
Figura 5.21	Resultados de la clasificación para la imagen de la figura 5.19 obtenidos mediante: (a) el clasificador BN y (b) el clasificador SO	188
Figura 5.22	Resultados de la clasificación para la imagen de la figura 5.19 obtenidos mediante: (a) el clasificador MF y (b) el clasificador MC	189
Figura 5.23	Imagen original perteneciente al conjunto S_3	190
Figura 5.24	Resultados de la clasificación para la imagen de la figura 5.23 obtenidos mediante: (a) el clasificador MF y (b) el clasificador MC	190

Índice de Tablas

Tabla 2.1	Ventajas e inconvenientes de los diferentes modelos de color	30
Tabla 3.1	Características generales de cada clasificador en las fases de aprendizaje y decisión	71
Tabla 3.2	Resumen de métodos de combinación y reglas de decisión	92
Tabla 3.3	Competencias relativas de cada clasificador	95
Tabla 4.1	Tabla de decisión normalizada y promediada según los criterios y sus pesos asociados	114
Tabla 4.2	Salidas de los clasificadores y densidades fuzzy asociadas	115
Tabla 5.1	Descripción de los conjuntos de datos utilizados para el ajuste de parámetros en la fase de entrenamiento	143
Tabla 5.2	Valores obtenidos para el parámetro h de la ecuación (5.1)	152
Tabla 5.3	Resumen de parámetros	157
Tabla 5.4	Comportamiento de los coeficientes $\hat{\Delta}_{BP}^{SE}(c)$ y $\hat{\Delta}_{BN}^{VC}(c)$ para S_1	163
Tabla 5.5	Estimación del número de clases para cada subconjunto S_i	164
Tabla 5.6	Variación de las funciones criterio $\alpha_{FC}^{SE}(\hat{c}_i)$ para cada clasificador m y para cada subconjunto S_i ; competencias absolutas promediadas de cada clasificador α_m ; competencias medias α^1 y α^2 para cada criterio CR_1 y CR_2 ; pesos específicos de cada criterio z_1 y z_2 ; densidades fuzzy g^m para cada clasificador	166
Tabla 5.7	Número de muestras y centros de las clases para los clasificadores individuales FC , BN , SO , $L1$ y $L2$ obtenidos a partir del conjunto de datos SI	167
Tabla 5.8	Matrices de covarianza para el estimador paramétrico BP según las clases	168
Tabla 5.9	Porcentajes medios de error y desviaciones estándar para cada PASO con los cuatro conjuntos de imágenes de prueba $S0$, $S1$, $S2$ y $S3$	176
Tabla 5.10	Número medio de iteraciones y tiempos medios de ejecución por imagen en cada uno de los conjuntos $S0$, $S1$, $S2$ y $S3$ en los tres PASOS	180
Tabla 5.11	Número de muestras, iteraciones y tiempos de ejecución para FC , SO , $L1$ y $L2$ durante la fase de entrenamiento	181

Capítulo 1

Introducción

1.1 Antecedentes

Es una evidencia palpable el hecho de que cada día existe un mayor número de sistemas equipados con sensores o dispositivos de captura de imágenes ofreciendo altas prestaciones, que de una u otra forma requieren un tratamiento adecuado de la información proporcionada con el fin de aprovechar al máximo su potencialidad. A continuación se citan algunos de tales sistemas para dar idea de los recursos disponibles.

1. Existen cámaras profesionales para aplicaciones técnico-industriales, que pueden llegar a proporcionar hasta 500 imágenes en color por segundo, en los manuales técnicos esta especificación aparece referenciada como 500 fps (“frames per second”), este es el caso de la cámara A504*kc* de Basler (2009), por citar sólo un ejemplo; esta prestación es una de las más elevadas, siendo las razones de captura más comunes del orden de entre 30 y 48 fps.
2. Cuanto mayor es la razón de captura menor suele ser la resolución espacial de las imágenes, es decir el número de píxeles por filas y columnas y viceversa (Pajares y Cruz, 2007*a, b*). Aunque no siempre esto es así, por ejemplo la cámara A404*kc* de Basler (2009) posee una razón de 96 fps con una resolución de 2352×1726 píxeles en las dimensiones horizontal (columnas) y vertical (filas) respectivamente. Por tanto, posee una mayor velocidad de captura y una mayor resolución que la cámara A504*kc* con 1280×1024 píxeles.

3. En el ámbito de la Teledetección existen por un lado los dispositivos de captura acoplados en satélites artificiales y por otro los sistemas instalados en vehículos aéreos tripulados.

En el primer caso, se incluyen algunos de los más famosos satélites comerciales entre los que se encuentran los siguientes (Satellite Image Corporation, SIC, 2009): *a)* GeoEye-1, lanzado en Septiembre de 2007 con resoluciones de 0.41 metros en la banda espectral del pancromático (PAN) y 0.25 metros en las bandas multiespectrales (MS, color), ver Pajares y col. (2007a) para los aspectos relacionados con las bandas espectrales. Para el año 2010/11 está previsto el lanzamiento del GeoEye-2 cuya resolución multiespectral desciende hasta los 0.25 metros; *b)* QuickBird, lanzado en Octubre de 2001, con resoluciones de 0.61 y 3.44 metros en el PAN y MS respectivamente, cubriendo áreas de aproximadamente $16.5 \times 16.5 \text{ Km}^2$; *b)* IKONOS, lanzado en Septiembre de 1999, con resoluciones de 1.0 y 4.0 metros en el PAN y MS respectivamente, con áreas de aproximadamente 11 a 13 Km^2 ; *c)* SPOT-5, lanzado en Mayo de 2002 y resoluciones de 5.0 y 10.0 metros en el PAN y MS respectivamente, con áreas de entre $60 \times 60 \text{ Km}^2$ a $80 \times 80 \text{ Km}^2$.

En el segundo caso, se encuentran las imágenes aéreas, que son precisamente las utilizadas en este trabajo. Las resoluciones espaciales y áreas de cobertura varían sustancialmente dependiendo de la altura de vuelo y los dispositivos de captura. Las imágenes aéreas de este trabajo, proporcionadas por Dimap (2009), poseen tamaños de hasta 33465×5817 píxeles con resoluciones espaciales de 0.25×0.25 metros en las direcciones horizontal y vertical respectivamente.

4. Otro de los sistemas de captura de imágenes es el dispositivo AVIRIS (2009), acrónimo de “Airborne Visible InfraRed Imaging Spectrometer”. Se trata de un único sensor que proporciona imágenes calibradas de hasta 224 canales o bandas espectrales contiguas, con longitudes de onda desde 400 a 2500 nanómetros. Este dispositivo se ha instalado en varios propulsores de la NASA (2009) entre los que destaca el turbo propulsor ER-2 volando aproximadamente a unos 20 Km de altura sobre la superficie terrestre.

5. Finalmente, debido al auge de los vehículos aéreos no tripulados (“UAVs: Unmanned Aerial Vehicles”) se ha suscitado un enorme interés por equiparlos con sistemas de captura de imágenes multibanda, incluyendo naturalmente las bandas ópticas y térmicas. Sin duda, su puesta en funcionamiento proporcionará gran cantidad de imágenes, que deberán ser convenientemente tratadas.

La información anterior proporcionada por los dispositivos mencionados debe adaptarse a las necesidades concretas de acuerdo con alguna aplicación de interés, en este sentido existen distintas áreas para su utilización, pudiéndose citar entre otras muchas las siguientes: Agricultura, Forestal, Conservación medioambiental, Meteorología, Detección de edificaciones o infraestructuras. Más concretamente, se puede precisar según las siguientes actuaciones:

1. Control agrícola, propiciado por la necesidad de conocer los cultivos programados para la recepción de subvenciones u otros aspectos relacionados.
2. Cómputo y medición de parcelas agrícolas y tipo de cultivos a los que se destinan.
3. Control de riegos agrícolas con el fin de evitar la sobreexplotación de acuíferos.
4. Agricultura de precisión para aplicar herbicida en el tratamiento de malas hierbas de forma selectiva, evitando la contaminación medioambiental y disminuyendo los costes de producción.
5. Prevención de posibles catástrofes naturales, mediante la identificación de zonas de alto riesgo: incendios, inundaciones, avalanchas, etc.
6. Evaluación de daños en catástrofes naturales: fuegos, daños por inundaciones, heladas en cultivos agrícolas, nevadas, etc.
7. Detección de cambios en determinadas zonas, principalmente urbanas para el control de edificaciones o impactos medioambientales agresivos.
8. Control de fenómenos geológicos tales como retrocesos de zonas de playa, erosión y sedimentación en determinadas áreas.
9. Vigilancia forestal o marítima.

10. Detección de infraestructuras o edificaciones: carreteras, caminos forestales, cañadas reales, etc.

De lo expuesto anteriormente se deduce la enorme cantidad de información que los sistemas de captura de imágenes pueden proporcionar. Surge, pues la necesidad de desarrollar procedimientos automáticos capaces de procesar esa ingente cantidad de datos de una forma adecuada, entendiendo por tal la obtención de resultados aceptables en el tiempo requerido dependiendo de la aplicación.

Dentro de las aplicaciones anteriormente mencionadas, la clasificación de imágenes según sus texturas, ha sido un campo muy estudiado en procesamiento de imágenes digitales desde hace bastante tiempo. Su objetivo consiste en realizar una partición de la imagen en varias regiones disjuntas, de tal forma que cada región se debe caracterizar por su homogeneidad con respecto a alguna medida previamente establecida. Este es el paso previo para abordar tareas dotadas con un mayor nivel de inteligencia y conocimiento como puede ser el reconocimiento de objetos, la comprensión de la imagen o la descripción de la escena (Gonzalez y Woods, 2007; Pajares y Cruz, 2007).

Si nos centramos en el campo de las imágenes de satélite, en Wilkinson (2005) se afirma que un importante porcentaje (sin cuantificar) de la investigación llevada a cabo en este ámbito está relacionada directa o indirectamente con la clasificación de texturas. Esta revisión se ha llevado a cabo en los quince años anteriores a la publicación, por lo que en el momento actual cabe realizar una extrapolación para llegar a conclusiones similares y por tanto más extensas.

Por mencionar una sencilla aplicación, en la figura 1.1 se muestra un ejemplo de clasificación cuya finalidad consiste en la identificación de las texturas naturales para realizar una planificación de posibles trayectorias a seguir por un vehículo terrestre de cara a una intervención ante la eventualidad de una posible catástrofe. Se trata de un ejemplo de simulación realizado por el grupo de investigación ISCAR (2006) dentro de sus trabajos de investigación con empresas del sector de la simulación, como es Indra Sistemas (2009).

En la figura 1.1(a) se muestra una imagen aérea original del terreno. En (b) la clasificación de texturas realizada mediante el estimador de máxima verosimilitud, que

se describe en el capítulo tres, sección 3.3.2. En (c) un camino óptimo (negro), encontrado mediante el algoritmo A* (Pajares y Santos, 2005), desde la base donde se encuentra el vehículo a un hipotético objetivo como consecuencia de la necesidad de actuación debida a una emergencia provocada por la aparición de un foco de fuego.

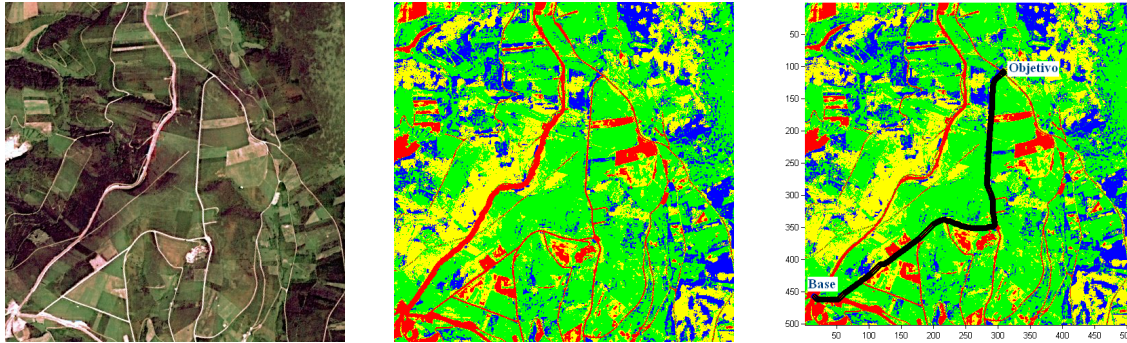


Figura 1.1 (a) Imagen aérea original; (b) identificación de las texturas; (c) camino óptimo

Fácilmente se deduce, de la aplicación anterior, que el tratamiento de las *texturas naturales* de las imágenes juega un papel primordial. Se plantea pues, el problema de la clasificación de las mismas por medio del mejor procedimiento posible. No en vano, siguiendo esta línea diversas empresas u organismos desarrollan o utilizan aplicaciones para abordar la problemática de la clasificación. Algunas de ellas son las que se mencionan a continuación:

1. Dimap (2009), dedicada en exclusiva al tratamiento de imágenes.
2. Proespacio (2009) agrupación de empresas del sector aeroespacial donde una de las actividades destacables es el desarrollo de aplicaciones mediante el uso de imágenes de satélite. En este consorcio destacan algunas empresas líderes del sector tanto en España como a nivel Internacional: EADS-Espacio, EADS Astrium, CRISA, GMV, Indra Espacio, Sener, Hispasat, IberEspacio, Inasmet, Insa, Mier, NTE, Tecnológica, Ryma, Hispasat, GTD, Alcatel, CRISA, GTD.
3. Organismos oficiales y centros de investigación con los que existen trabajos de colaboración previos por parte de miembros del grupo ISCAR (2006)
 - Instituto de Automática Industrial (IAI) del Consejo Superior de Investigaciones Científicas (CSIC) (Pajares y col., 2007; Tellaache y col., 2008)
 - Instituto Nacional de Técnica Aeroespacial (INTA) (Pajares, 2006)

- Centro de Estudios y Experimentación de Obras Públicas (CEDEX), (Pajares y col. 2001, Pajares y col. 2002)

La mayoría de las empresas citadas y algún organismo también mencionado, utilizan para desarrollar sus aplicaciones herramientas comerciales tales como las relacionadas a continuación:

- ERDAS Imagine (2009)
- Paquete de productos de información geográfica de Intergraph (2009)
- ENVI (2009) en lenguaje IDL (Canty, 2006)
- PCI Geomatics (2009)
- E-Cognition y Definiens (2009)
- Image Processing Toolbox de MATLAB (2009)

En general dichas herramientas poseen módulos de clasificación supervisada, de forma que requieren la intervención del usuario, quien ha de seleccionar y suministrar las muestras necesarias para el entrenamiento. Estos módulos suelen estar basados en clasificadores clásicos entre los que se encuentran los relacionados a continuación (Pajares y Cruz, 2007):

- Clasificadores estadísticos Bayesianos: máxima verosimilitud, ventana de Parzen.
- Agrupamiento borroso o fuzzy (“Fuzzy Clustering”).
- Algoritmos competitivos: Cuantización vectorial (“Learning Vector Quantization”), Mapas Auto-organizativos (Self-Organizing Feature Maps).
- Red neuronal retropropagación (“backpropagation”).

Excepto la red neuronal, los clasificadores anteriores se describen en el capítulo tres por ser la base para la combinación de los métodos propuestos en este trabajo de investigación.

A título informativo, tanto la herramienta ENVI como E-Cognition incorporan métodos de decisión basados en reglas, por lo que han abierto una línea de aplicación centrada en los sistemas de producción dentro del campo de la Inteligencia Artificial (Pajares y Santos, 2005).

Aunque bien es cierta la evidente mayor potencialidad de las herramientas citadas, no es menos cierto que los retos tecnológicos derivados de las aplicaciones mencionadas previamente, hacen que en algunos casos la utilización de tales herramientas sea insuficiente para abordar las propuestas de proyectos demandados por los clientes según sus necesidades reales. E incluso, aún siendo suficientes siempre cabe la posibilidad de introducir mejoras o modificaciones precisas según dichas necesidades. Esto es lo que ocurre realmente en numerosas aplicaciones. Sobre esta base de razonamiento, la *clasificación* de texturas naturales se perfila claramente como uno de los procedimientos susceptibles de mejora, esto es así porque en la mayoría de los casos, los métodos simples de clasificación no proporcionan los resultados esperados, particularmente en imágenes de exterior como las utilizadas en este trabajo, cuya base son las texturas naturales. De aquí, surge la necesidad de investigar nuevas alternativas en el ámbito de la clasificación, orientada hacia la identificación de las mencionadas texturas naturales.

1.2 Identificación del problema

Los clasificadores referidos previamente han sido probados, bien a través de las herramientas citadas o en experimentos descritos en la literatura, siendo una de las aplicaciones de prueba la clasificación de texturas naturales en dichas imágenes de exterior, como se expone en el capítulo dos.

Existen dos problemas claramente identificados en este tipo de clasificaciones. Uno es el que se deriva de la naturaleza supervisada misma de los clasificadores, ya que requieren la presencia del supervisor, que generalmente es el usuario. El segundo y más importante, desde el punto de vista de este trabajo, es que los resultados de la clasificación obtenida por dichas estrategias son claramente mejorables.

Para resolver ambos problemas, se ha intentado la búsqueda de soluciones con mayor o menor éxito. Así, para evitar la presencia del supervisor y conseguir de este modo un método automático, se han propuesto métodos, que determinan la mejor partición o clasificación de los datos a base de realizar varias pruebas hasta conseguir un resultado satisfactorio (Duda y col., 2001). Respecto de la búsqueda de estrategias para la mejora de los resultados, la tendencia actual, según se desprende de la revisión bibliográfica realizada en el capítulo dos, se inclina por la combinación de clasificadores bajo la hipótesis de que la suma de esfuerzos suele mejorar los resultados individuales, como en

tantos otros aspectos de la vida real. En esta línea, la comunidad científica viene realizando un importante esfuerzo desde hace ya más de una década (Kuncheva, 2004; Kittler y col., 1998).

La clasificación de texturas en imágenes naturales no es ajena, ni mucho menos, a la problemática general expuesta hasta aquí. En efecto, cuando uno se enfrenta a una imagen natural como la de la figura 1.2, observa la existencia de una serie de texturas claramente identificadas, como puede ser el caso de zonas de campos de cultivo, zonas boscosas, edificaciones, caminos, carreteras, etc. Ahora bien, descendiendo a un nivel de mayor detalle se observa que dentro de una determinada textura, por ejemplo los campos de cultivo, aún existe una amplia diversidad de matices en las texturas. Esto mismo ocurre en zonas boscosas donde la abundancia de matorrales diferencia unas zonas de otras de similares características, e incluso la misma observación aflora también en el caso de las texturas de zonas urbanas. La pregunta que se puede plantear aquí es ¿cuál es el número de texturas a identificar? La respuesta a esta pregunta no es precisamente obvia, para ello se requiere establecer un método que determine el mejor número de texturas existentes. Estamos por tanto, ante uno de los objetivos de este trabajo de investigación.



Figura 1.2 Diversidad de texturas naturales

Por otra parte, si la comunidad científica propone la combinación de clasificadores como la estrategia más apropiada para lograr una mejora en la clasificación, es lógico pensar en seguir esta línea. Si bien, en este sentido surge una nueva cuestión, ¿cuál es la mejor de todas las combinaciones posibles? Es evidente que no se podrán probar todas las

combinaciones factibles debido al elevado número de combinaciones que se obtiene. Ahora bien, gracias a la investigación bibliográfica realizada, se abre una nueva puerta por la cual se pueden mejorar los resultados de la clasificación mediante la combinación de clasificadores. Esto constituye el segundo y fundamental objetivo de este trabajo.

1.3 Motivación y Objetivos

1.3.1 Motivación

Desde el punto de vista de su utilidad industrial, el trabajo de investigación desarrollado tiene su origen en la necesidad planteada por la empresa Dimap (2009) con el fin de desarrollar una serie de trabajos para el Servicio Territorial de Galicia (SITGA) encaminados al análisis de texturas mediante imágenes aéreas correspondientes a la región de Abadín en Lugo (Galicia, España). En estos trabajos la identificación y clasificación de las texturas constituye la base de los mismos, siendo el objetivo final disponer de esta información para facilitar las tareas encaminadas a la ordenación del territorio, además de identificar aquellas zonas de riesgo de cara a la realización de tareas de prevención de incendios. Las imágenes utilizadas en los experimentos son cortesía de ambas entidades, quienes nos las han proporcionado sin impedimento alguno.

Por otro lado, el grupo ISCAR (2006) tiene abierta una línea de investigación en tareas relativas a la clasificación de texturas naturales, como consecuencia del desarrollo de varios proyectos donde dichas tareas se incluyen como fundamentales. La relación entre esas tareas y la investigación desarrollada en este trabajo surge como consecuencia de la colaboración en las mismas en virtud de los correspondientes contratos suscritos entre la Universidad Complutense a través de su Oficina de Transferencia Tecnológica y las respectivas empresas o entidades. También por la participación, como miembro del equipo de investigación, en el caso de los proyectos del Plan Nacional de I+D. A continuación se detallan las líneas generales de tales proyectos, incidiendo en los objetivos relacionados con la identificación de las texturas naturales:

1. Proyecto titulado *Análisis estereoscópico y reconocimiento de texturas en imágenes forestales obtenidas mediante el dispositivo de medición MU-200501738* al amparo del artículo 83 de la LOU (Ley Orgánica de Universidades). El origen de este proyecto se encuentra en el interés suscitado por el Centro de

Investigación Forestal (CIFOR) del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) para automatizar el proceso de extracción de información relativa al volumen de madera existente en un área forestal mediante el dispositivo con número de patente MU-200501738. Se trata de un sistema de visión estereoscópica con el que se pretende determinar las distancias a las que se encuentran los árboles y a partir de ahí su altura y volumen de madera. Para establecer las correspondencias estereoscópicas es necesario identificar las texturas existentes en la imagen con el fin de clasificarlas, bien como troncos o ramas de los árboles, que constituyen los objetos de interés, o bien como otros tipos de texturas tales como suelo, cielo, hojas de árbol, matorrales o hierbas, que deben ser excluidas por carecer de interés e incluso perjudicar el proceso de correspondencia. Este proyecto en la actualidad está en fase de ampliación por parte de las entidades involucradas.

2. Durante el trienio 2006-2008 se ha colaborado en el subproyecto CICYT titulado *Extracción de Conocimiento Espacio-Temporal y Visión Artificial para la Automatización de Tratamientos Localizados* (AGL2005-06180-C03-03) con el Instituto de Automática Industrial del CSIC. Este proyecto tiene su continuidad en un nuevo proyecto del Plan Nacional de I+D 2008 para el trienio 2009/2011 concedido con el número de referencia AGL2008-04670-C03-02/AGR, con el título *PERcepción vIsual, toma de deCisión y Actuación en sistemas artificiales complejos: control selectivo de malas hierbas en cultivos agrícolas (PERICIA)*. En ambos proyectos la identificación de las texturas existentes en las imágenes (cultivo, malas hierbas, suelo) es una tarea primordial para localizar espacialmente los rodales de malas hierbas en campos de cultivo de cereal y maíz con el fin de proceder a su pulverización selectiva. El objetivo final del proyecto consiste en minimizar el efecto contaminante medioambiental por el tratamiento con herbicidas de los campos de cultivo, así como disminuir los costes de producción al necesitar una menor cantidad de producto.
3. Con la empresa TCP Sistemas e Ingeniería se participa en dos proyectos titulados respectivamente *AUTOROVER: estudio de autonomía basada en imágenes para “rover” de exploración planetaria* y *Visión estereoscópica para Auto-rover: estudio de autonomía basada en imágenes*. El primero procede de la participación

de la citada empresa en la convocatoria pública según la orden 2259/2007 (B.O.C.M 272 de 15/11/2007) para el fomento de la innovación en el sector aeroespacial de la Comunidad de Madrid con financiación del Fondo Europeo Regional (Referencia 04-AEC0800-000035/2008). El segundo corresponde a una ampliación del anterior dentro del programa Nacional de Proyectos de Investigación en el Sector Aeroespacial según la orden PRE/998/2008 (BOE 11/04/2008) del Ministerio de la Presidencia (Referencia SAE-20081093). Ambos proyectos tienen como fundamento la navegación autónoma de robots en la superficie de Marte e incluyen como uno de los objetivos clave, el análisis de las texturas observadas a través del sistema de visión en dicha superficie. El fin último consiste en discernir entre diferentes tipos de estructuras con el objetivo de determinar la navegabilidad de dichas superficies para los vehículos autónomos.

Además, desde el punto de vista tecnológico, la motivación viene suscitada por el hecho de tratar de mejorar los resultados de clasificación considerando la problemática expuesta previamente. Desde esta perspectiva, surgen los siguientes planteamientos:

1. Interés por diseñar un procedimiento de clasificación automático utilizando técnicas de validación cuantitativa, en contraposición a las cualitativas según el criterio del experto humano. Todo ello con la finalidad de conseguir un método de naturaleza *no supervisada* con vistas a la automatización del proceso.
2. Simultáneamente, surge la idea de investigar procedimientos combinados de clasificadores lo más eficientes posibles. Siguiendo esta línea, se propone el desarrollo de clasificadores de naturaleza local, que consideran cada píxel de la imagen como una entidad en sí misma, que debe ser clasificada.
3. Finalmente, debido a que en las imágenes se mantienen las relaciones espaciales de la escena, se pueden observar dos aspectos de relevancia relacionados con un píxel dado y sus vecinos, según lo cual siempre es posible el diseño de métodos combinados. Una primera observación sugiere que un píxel perteneciente a una determinada región está probablemente rodeado de píxeles de la misma textura si se trata de un píxel interno a la región definida por la textura. La segunda observación proviene de la consideración de aquellos píxeles que hacen frontera entre dos o más regiones con diferentes texturas, lo que conduce a que en la

vecindad existen píxeles que pertenecen a más de una textura. En cualquier caso, estas observaciones sugieren la posibilidad de que se consideren ambas observaciones durante el proceso de clasificación. De aquí surge el interés por investigar técnicas de combinación de clasificadores de naturaleza global, entendiendo por tal las relaciones espaciales entre un píxel y sus vecinos. Esto constituye una aportación relevante del presente trabajo de investigación.

En definitiva, la idea consiste en desarrollar técnicas automáticas de clasificación tanto de naturaleza local como global, derivadas respectivamente de la primera y segunda observación.

1.3.2 Objetivos

A la vista de las consideraciones expresadas en los apartados precedentes, se plantean los siguientes objetivos de investigación:

1. Estudio bibliográfico de la literatura existente a nivel de clasificación de texturas en imágenes.
2. Diseñar estrategias de clasificación no supervisada.
3. Diseñar métodos eficientes de combinación de clasificadores a nivel local.
4. Diseñar métodos de combinación de clasificadores a nivel global.
5. Analizar los resultados extrayendo las pautas necesarias para la combinación de clasificadores.
6. Identificar las líneas de investigación futuras.
7. Divulgar y difundir los resultados de la investigación mediante publicaciones científicas u otros mecanismos.

1.4 Metodología

En el desarrollo del trabajo se han seguido las directrices marcadas por el método científico. Inicialmente, se ha realizado una búsqueda de información bibliográfica en todas las fuentes disponibles, utilizando principalmente para ello el acceso a través de la red del Consorcio Madroño al que están suscritas las Universidades de la Comunidad de Madrid. Se han estudiado los antecedentes metodológicos y se ha analizado y evaluado el

estado de la ciencia en este área. Seguidamente, se han dado los pasos que se indican a continuación, a la vez que se han revisado de manera continua las fuentes bibliográficas hasta el final de la investigación e incluso en el momento de la redacción de este trabajo. Las etapas seguidas han sido:

1. Definición del problema de investigación a raíz tanto de las necesidades identificadas como de la motivación establecida previamente.
2. Planteamiento de los objetivos de la investigación.
3. Formulación de planteamientos verificables.
4. Obtención de datos experimentales.
5. Clasificación, análisis e interpretación de los datos.
6. Conclusiones generales.
7. Descripción de líneas futuras de investigación.

Para la realización de los experimentos se han creado los sistemas de clasificación basados en las técnicas propuestas. Este proceso abarca desde la adquisición de datos para la creación de una base suficientemente amplia con la que experimentar hasta la obtención de resultados, pasando por su visualización.

1.5 Aportaciones de la investigación

A partir de los objetivos propuestos y teniendo en cuenta los aspectos destacados previamente, se trata de dar solución a los problemas planteados y de paso facilitar a la comunidad científica una serie de estrategias de solución, que pueden extenderse a otros problemas de similar naturaleza. Las contribuciones realizadas por este trabajo de investigación se sintetizan en los siguientes puntos:

1. Se propone un procedimiento de clasificación no supervisada a partir del método de agrupamiento fuzzy o borroso supervisado. En este trabajo se utilizarán los términos fuzzy y borroso indistintamente, sin ningún matiz diferenciador.
2. En cuanto a los procedimientos basados en la combinación de clasificadores, se proponen dos estrategias de naturaleza local y otras dos de naturaleza global. En ambos casos, se proporcionan las pautas necesarias para su aplicación a otros

ámbitos según las líneas de actuación siguientes: *a)* mediante la incorporación de nuevos clasificadores y *b)* mediante el establecimiento de las bases y sus posibilidades de aplicación en otros contextos de clasificación diferentes a los de las texturas.

Las aportaciones y divulgación de resultados, hasta el momento actual, se recogen en los siguientes trabajos:

1. En Guijarro y Pajares (2009) se propone un esquema de clasificación no supervisada a partir del clasificador conocido como Agrupamiento Borroso de naturaleza supervisada. En su conjunto, se trata de un método de índole local, de suerte que cada píxel se clasifica atendiendo a los criterios proporcionados por seis clasificadores individuales. Este procedimiento se describe en la sección 4.2 donde la combinación se lleva a cabo bajo el paradigma de la teoría de la decisión *multi-criterio fuzzy*. Este trabajo se encuentra publicado en la revista *Expert Systems with Applications* que publica diez números por año con un índice de impacto para el año 2007, último publicado, de 1.177.
2. En Guijarro y col. (2008) se propone igualmente una estrategia de naturaleza no supervisada con el mismo esquema de la anterior. En este caso se combinan dos clasificadores individuales bajo una perspectiva global utilizando la información de la vecindad de un píxel en la imagen. El proceso global consiste en un procedimiento de relajación iterativo hasta conseguir el máximo grado de estabilización. Se trata del método descrito en la sección 4.4 y publicado en la serie *Lecture Notes in Computer Science* bajo la subserie *Lecture Notes in Artificial Intelligence*.
3. En Guijarro y col. (2007a) el procedimiento propuesto es de naturaleza local y consiste en la combinación del método de agrupamiento borroso con el método no paramétrico de la ventana de Parzen, descritos respectivamente en las secciones 3.3.1 y 3.3.3. La mencionada combinación se lleva a cabo bajo el paradigma de la probabilidad condicionada de Bayes. Mediante el clasificador de agrupamiento borroso se obtiene la probabilidad a priori, mientras que el estimador de la ventana de Parzen proporciona la probabilidad a posteriori. El método es de naturaleza no supervisada basado en un esquema similar a los

mencionados en los puntos anteriores. Este trabajo se encuentra publicado en las actas del Simposio auspiciado por la sociedad IEEE, según la citada referencia.

4. En Guijarro y col. (2007b) se presenta una estrategia de naturaleza local, que utiliza para la combinación el método de agrupamiento borroso y el método de cuantización vectorial, descritos en las secciones 3.3.1 y 3.3.6 respectivamente. El planteamiento vuelve a ser de naturaleza no supervisada, de forma que el primero establece una partición inicial con los datos disponibles, la cual es transferida al segundo y mejorada hasta conseguir su validación. Este método está publicado en la serie Lecture Notes in Computer Science bajo la subserie Lecture Notes in Artificial Intelligence.
5. En Guijarro y col. (2007c) se establece un planteamiento local similar al propuesto en el punto cuatro con la diferencia de que en lugar del método de agrupamiento borroso se utiliza el estimador de máxima verosimilitud descrito en la sección 3.3.2. El esquema no supervisado sigue la misma estrategia que en el caso del trabajo del punto cuatro. Este método está publicado en las actas del II Congreso Español de Informática, según los detalles de la referencia.

Los métodos descritos en los puntos tres a cinco han quedado superados por el método descrito en el punto uno, tanto desde el punto de vista de su planteamiento como en cuanto a los resultados obtenidos, razón por la cual no se describen en esta memoria.

1.6 Organización de la tesis

La memoria de tesis se presenta estructurada en capítulos. Estos capítulos se organizan siguiendo el orden natural de la investigación, cuya distribución es la que se ofrece a continuación:

1. Primer capítulo. *Introducción*, se estudian y plantean los objetivos de la tesis, así como la motivación e introducción de las propuestas y planteamientos realizados, tal y como se ha expuesto en las secciones precedentes.
2. Segundo capítulo. *Estado del arte*, donde se revisan diferentes estrategias existentes en la literatura, que proporcionan los argumentos y la base de la investigación.

3. Tercer capítulo. *Clasificadores clásicos: simples y combinados*, se describen, desde una perspectiva teórica, los métodos de clasificación que tienen que ver con la propuesta de investigación realizada en la tesis. Así, los métodos individuales constituyen la base para la combinación de las estrategias planteadas, mientras que los combinados sirven como referencia para la comparación y comprobación de la eficacia de los métodos propuestos.
4. Cuarto capítulo. *Estrategias de combinación de clasificadores*, se exponen los esquemas de clasificación combinados, que constituyen la aportación esencial de este trabajo de investigación.
5. Quinto capítulo. *Análisis de resultados*, se presentan los resultados obtenidos y se analizan comparativamente con otros métodos individuales y combinados existentes con anterioridad en la literatura. Se establece una discusión y valoración de los mismos.
6. Sexto capítulo. *Conclusiones*, se exponen las líneas generales seguidas, realizando un balance general del trabajo. Se plantean las líneas de investigación futuras así como posibles mejoras o extensiones de los métodos propuestos.

Capítulo 2

Estado del arte

2.1 Visión general

En este capítulo se recopila información sobre la utilización de combinaciones de clasificadores aplicados a imágenes digitales; a veces, a este tipo de clasificadores se les denomina híbridos en la literatura especializada. En cualquier caso, de forma general, un clasificador combinado tiene su fundamento en la participación de otros clasificadores individuales, que constituyen las unidades básicas de la combinación. Se analizan los aspectos relacionados con la combinación de clasificadores desde las perspectivas más comúnmente utilizadas en la literatura, orientándose hacia la clasificación de texturas en imágenes naturales de exterior, que constituyen el objetivo de la investigación. Los aspectos tratados se estructuran como sigue:

- 1) En primer lugar, se realiza un recorrido por la literatura para identificar métodos generales de clasificación aplicados a imágenes, sección 2.2.
- 2) En segundo lugar se realiza un estudio relativo a la conveniencia de utilizar métodos de clasificación combinados, sección 2.3.
- 3) En tercer lugar, se estudian métodos cuyo fundamento son los píxeles como unidades básicas de las imágenes o agrupaciones de los mismos según ciertas distribuciones regulares. Por tanto, en esta categoría se incluyen tanto las propiedades espectrales (color), como espaciales (distribuciones de texturas) o ambas bajo un enfoque combinado, sección 2.4.

- 4) En cuarto lugar, se analizan específicamente métodos cuyo fundamento son las relaciones de vecindad existentes entre un píxel y los que le rodean. Se contemplan las agrupaciones irregulares espaciales de píxeles, que son realmente como aparecen en las imágenes naturales objeto de este trabajo de investigación, sección 2.5.
- 5) Finalmente, en la sección 2.6, se resumen las diferentes técnicas analizadas.

La mayoría de los métodos existentes en la literatura utilizan más de una estrategia o método de clasificación con diferentes características. Por tanto, no resulta fácil ubicar un determinado trabajo de investigación exclusivamente en un epígrafe, razón por la cual algunos de los métodos expuestos aparecen citados en varias secciones; no obstante se ha procurado encuadrar cada uno de ellos según la aportación más relevante de acuerdo con el contenido de la sección.

2.2 Métodos de clasificación de texturas en imágenes

Durante las últimas tres décadas e incluso cuatro, se han propuesto en la literatura diversos métodos encaminados a la segmentación de las imágenes según sus texturas, ver por ejemplo Xia y col. (2007) y referencias asociadas. En Wilkinson (2005) se realiza una exhaustiva revisión de métodos de clasificación durante los quince años previos a la publicación de dicho trabajo aplicados a las imágenes de satélite.

La mayoría de ellos pueden considerarse como métodos con dos procesos sucesivos asociados: *extracción de características* y *segmentación basada en las características*. El término sucesivos hace referencia al hecho de que se realizan en ese orden durante la clasificación.

La extracción de características tiende a encontrar una medida apropiada para caracterizar la homogeneidad de cada región dentro de la imagen, estableciendo una asociación entre la región y una determinada textura (Valdovinos y col. 2005, Puig y García, 2006 o Hanmandlu y col. 2004). Algunas de tales características se analizan posteriormente en este mismo capítulo en la sección 2.4.

La segmentación basada en las características consiste en el proceso de asignar a cada píxel de la imagen una etiqueta que lo identifique como perteneciente a una determinada

clase. Además, se han diseñado y utilizado una amplia gama de algoritmos tanto de naturaleza supervisada como no supervisada. A continuación se realiza un esbozo de los más relevantes, si bien algunos de ellos serán estudiados con mayor profundidad en el capítulo tres, dado que constituyen la base de la propuesta de combinación realizada en este trabajo.

1. Técnicas borrosas o fuzzy

Asumen el hecho de que los colores tienden a agruparse en la imagen formando clases que determinan la distinta naturaleza de las texturas. Dada la naturaleza fuzzy, en cada una de las clases, se admiten píxeles con cierta ambigüedad, en el sentido de que un mismo píxel podría pertenecer tanto a una textura como a otra según sus características. Este hecho se manifiesta más claramente en las zonas fronterizas de las texturas, donde se producen transiciones graduales en los valores de los píxeles de una zona a otra de la frontera. A la categoría fuzzy pertenece uno de los métodos más clásicos en clasificación como es el conocido Agrupamiento borroso o en inglés “Fuzzy Clustering”. Para la fase de entrenamiento utiliza un método de optimización, que permite conseguir agrupaciones compactas y resolver muchas de tales ambigüedades. En esta línea se encuentran los trabajos de Pajares y Cruz (2002), Pajares y col. (2002) o Lim y Lee (1990) entre otros.

Mohaddamzadeh y Bourbakis (1997) definen las regiones a segmentar como agrupaciones de píxeles que poseen similar color, mientras los cambios de región se manifiestan gradualmente como ya se ha indicado previamente. Para plasmar esta idea definen una función de naturaleza fuzzy en el espacio de color RGB (Pajares y Cruz, 2007a).

Cheng y col. (2001) utilizan el concepto de “homogram” para extraer regiones homogéneas en una imagen de color; a su vez definen el concepto de homogeneidad fuzzy para extraer regiones de similares características mediante la obtención de una serie de umbrales. Tras la segmentación vuelven a aplicar un criterio de similitud para fusionar regiones, se trata de un proceso similar al método de segmentación de regiones según la técnica de árboles cuaternarios (Pajares y Cruz, 2007a).

En el ámbito de las imágenes de satélite los siguientes trabajos utilizan métodos de clasificación basados en la lógica fuzzy: Seong y Usery (2001), Zhang y Foody (1998).

2. *Redes neuronales*

Por el carácter no lineal, que permite una mayor separabilidad en el espacio de características y su alta capacidad para desarrollar procesamiento en paralelo, las redes neuronales se han utilizado también en la clasificación de texturas en imágenes en general y en imágenes de satélite en particular (Mas y Flores, 2008).

Dentro de las redes neuronales se encuentra la red neuronal de Hopfield (Haykin, 1994), compuesta por una serie de neuronas o nodos conectados entre sí. Este tipo de red realiza un proceso de optimización basándose en la minimización de una función de energía hasta que la propia red adquiere un grado de estabilidad aceptable. En Huang (1999) se detectan los picos más significativos del histograma y se diseñan tres redes de Hopfield, una para cada banda de color, donde los nodos de cada red son los píxeles de la correspondiente banda espectral y las conexiones establecen similitudes entre ellos. En Campadelli y col. (1997) se diseña un algoritmo con $M \times N \times S$ neuronas siendo M y N las dimensiones de la imagen y S el número de clases que se obtienen mediante el análisis previo del histograma. En este tipo de redes se determinan las conexiones entre los nodos mediante el cómputo de funciones de consistencia entre los píxeles, basándose en las etiquetas de los vecinos. Para ello, se requiere la realización de un proceso de clasificación y etiquetado previo.

La red retro-propagación ha sido utilizada en clasificación de imágenes naturales, incluyendo las imágenes de satélite, cabe mencionar entre otros los trabajos de Pajares y col. (2001), Atkinson y Tatnall (1997) o Paola and Schowengerdt (1995), donde las componentes espectrales correspondientes a los tres colores básicos de cada píxel constituyen las entradas a la red, mientras que como salidas se establecen las clases a las que pertenecen dichos píxeles según un mecanismo de aprendizaje supervisado.

Los mapas Auto-Organizativos (Self Organizing feature Maps, SOM) han sido otra de las herramientas utilizadas en clasificación. Este es el caso de los trabajos desarrollados en Jiji y Ganesan (2008) o Lo y Pei (1999) entre otros. En ellos los centros de las clases se estiman según la correspondiente ley de aprendizaje. Este método se describe exhaustivamente en el capítulo tres, sección 3.3.4, debido a que se trata de uno de los clasificadores que participa en dos de las estrategias combinadas propuestas en este trabajo de investigación.

3. *Cuantización vectorial (“Learning Vector Quantization”)*

Se trata de un procedimiento que puede ser de naturaleza supervisada o no supervisada. En su versión supervisada se determina de antemano el número de clases existentes en la imagen, mientras que en la no supervisada es a través del propio método como se establece el número de clases en función de un umbral fijado previamente. En cualquiera de las dos versiones, se trata de un proceso de aprendizaje, que obtiene los centros de las clases. Algunos trabajos basados en estos métodos, que además utilizan información del color, son los propuestos en Giacinto y col. (2000), Chan y col. (2003) o del Frate y col. (2007). Los trabajos de Lillo y col. (2007), Aiyer y col. (2005) o Pyun y col., (2002) realizan una cuantización vectorial utilizando propiedades de texturas obtenidas mediante la aplicación de la transformada de Fourier (Pajares y col. 2007a; Gonzalez y Woods, 2007). La versión supervisada de este método se describe en el capítulo tres, sección 3.3.5, por su participación en las estrategias combinadas propuestas en este trabajo de investigación.

4. *Vecinos más próximos*

Este método se conoce en la literatura inglesa como k -Nearest Neighbours (k -NN). En ellos, dado un conjunto de clases, los nuevos píxeles a clasificar con sus características, se asignan a aquella clase que posea un número de k vecinos más próximos. Dicha proximidad se determina mediante alguna medida de similitud, por ejemplo una distancia. En Haapanen y col. (2004) se utiliza esta técnica para la segmentación de zonas boscosas en imágenes aéreas. En Muñoz-Marí y col. (2007) se utiliza este clasificador como referencia para el análisis comparativo de su método. En Singh y col. (2001) se aplica este método para clasificar nueve tipos diferentes de texturas.

5. *Estimas de máxima verosimilitud*

Consistentes en la estimación de una función de densidad de probabilidad, que constituye el objeto del aprendizaje. Tomando como referencia la función de densidad de probabilidad estimada, es posible determinar la probabilidad de pertenencia de una muestra dada a las diferentes clases. En esta línea se encuentran los trabajos de Frizelle y Moody (2001), Newsam y col. (2004), Xie y Tsui (2004), Trauwaert y col. (1991) o

Benediktsson y col. (1990). Muñoz-Marí y col. (2007) utilizan también este clasificador para comparar los resultados obtenidos con su método.

6. *Algoritmos genéticos*

Utilizados como métodos de optimización que aprovechan la información espacial de la vecindad para definir las funciones objetivo, tal es el caso de Tso y Mather (1999).

7. *Máquinas de vectores soporte*

Estas técnicas están especialmente diseñadas para aplicaciones bi-clase, si bien se han propuesto ampliaciones para abordar el problema multi-clase. Tal es el caso de los trabajos propuestos en Camps-Valls y Bruzzone (2005), Camps-Valls y col. (2004) o Melgani y Bruzzone (2004). En Ferreiro-Armán (2007) se realiza un estudio sobre el comportamiento en imágenes de satélite de las máquinas de vectores soporte, además del Agrupamiento borroso, los Mapas Auto-organizativos y la red neuronal retro-propagación.

8. *Árboles de decisión*

Se presentan como una alternativa a los métodos clásicos de clasificación donde lo que se infieren son reglas de decisión, que permiten realizar las clasificaciones en base al cumplimiento de las premisas aprendidas y reflejadas en las reglas (Tottup, 2007; Kumar y Majunder, 2001; Hansen y col. 1996). En Debeir y col. (2002) se utiliza el esquema C4.5 para la clasificación de imágenes aéreas.

Un aspecto importante a destacar respecto de los métodos mencionados en los puntos uno a ocho, es el hecho de que se centran exclusivamente en la clasificación de los píxeles sin tener en cuenta las características espaciales de las imágenes. Así, para cada píxel de la imagen se extraen las correspondientes propiedades o características, procediendo seguidamente a su clasificación sin tener en cuenta su posición o relación con otros píxeles vecinos en la imagen. Con el fin de incorporar la información espacial, se han realizado diferentes esfuerzos para considerar esta importante propiedad de las imágenes. En efecto, una imagen se define como una estructura rectangular donde cada píxel se ubica en una determinada localización espacial dentro de la imagen. Después de extraer las características de cada píxel, la imagen puede expresarse como agrupaciones

de características, de forma que cada agrupación define una región. Cada región posee características similares, además de una cierta compacidad espacial y probablemente, constituya una parte de un objeto, por ejemplo una textura. De aquí surge la idea de que durante el proceso de clasificación no sólo se deben tener en cuenta las características individuales de los píxeles sino también su localización e información espacial. En este sentido, se han propuesto diferentes estrategias para incorporar la información espacial en los métodos de clasificación. Posteriormente, en la sección 2.5 de este mismo capítulo, se aborda específicamente este aspecto.

2.3 Métodos de clasificación combinados

Una importante línea de investigación actual para mejorar los resultados de la clasificación obtenidos mediante clasificadores individuales, consiste en combinar éstos. La combinación de clasificadores individuales constituye el objetivo de esta sección.

En primer lugar, se abordan los aspectos relacionados con la combinación desde una perspectiva general, para posteriormente tratar de forma específica la combinación para la segmentación de las texturas en imágenes.

2.3.1 Clasificadores combinados de índole general

Un aspecto importante a destacar, asumido por la comunidad científica en general, es el hecho de que mediante la combinación de clasificadores se obtienen mejores resultados que con los clasificadores utilizados de forma individual, esto es lo que se concluye en los trabajos de Valdovinos y Sánchez (2007), Valdovinos y col. (2005), Kuncheva (2003, 2004), Kumar y col. (2002), Kittler y col. (1996, 1998) o Cao y col. (1995) entre otros muchos. Por otro lado, además existen diferentes estudios relativos al tema de la combinación, en ellos se destacan las ventajas de este planteamiento en el sentido de que la combinación permite resaltar las bondades de los clasificadores al mismo tiempo que se atenúan o desaparecen las desventajas (Partridge y Griffith, 2002; Deng y Zhang 2006). Según los resultados obtenidos en los trabajos citados, la justificación sobre la conveniencia de utilizar clasificadores queda suficientemente justificada. Surge, en este momento, la cuestión de cómo llevarla a la práctica. En este sentido, destacan dos importantes estrategias, comúnmente aceptadas por la comunidad científica, para llevar a

cabo la combinación de clasificadores, a saber (Valdovinos y col. 2005): *selección* y *fusión*.

En la selección, cada clasificador individual realiza una decisión y se elige un único clasificador como el mejor de acuerdo a algún criterio preestablecido. Por el contrario, en la fusión, todos los clasificadores participan de una u otra forma según la estrategia de combinación utilizada. Tomando como referencia los trabajos de Kuncheva (2003, 2004), Grim y col. 2002, Kittler y col. (1998, 1996) o Duda y col. (2001), en el capítulo tres, sección 3.5 se describen detalladamente ambos enfoques por constituir la base de comparación para los métodos propuestos en este trabajo de investigación.

En cualquier caso, existen algunas estrategias donde la combinación se realiza mediante una o más redes neuronales, que reciben como entradas las salidas de los clasificadores individuales con los que realizan el correspondiente proceso de aprendizaje. En esta línea se encuentran los trabajos de Wozniak y Jackowski (2008), Hashem y Schmeiser (1995) o Hansen y Salamon (1990).

2.3.2 Combinación de clasificadores en texturas de imágenes

En el ámbito de la clasificación de texturas naturales en imágenes, se pueden mencionar algunos trabajos de interés que han utilizado combinaciones de algunas de las técnicas anteriormente expuestas. En este sentido se encuentran los trabajos de Lepisto y col. (2003, 2005), quienes proponen un nuevo método de clasificación donde las salidas de los clasificadores individuales forman los vectores patrón para el nuevo proceso de clasificación. Dichos vectores contienen las opiniones de los clasificadores individuales en cada una de sus componentes. De esta manera, la decisión se lleva a cabo comparando estas opiniones con respecto a las almacenadas para las muestras y utilizando un criterio de mínima distancia tal como la distancia de Hamming. Este procedimiento ha sido comparado favorablemente frente a los clasificadores combinados clásicos tales como la regla del máximo, media o el método conocido como votación mayoritaria, propuestos en Kittler y col. (1996, 1998) y descritos en la sección 3.5.

Muñoz-Marí y col. (2007) proponen un método combinando los cuatro clasificadores siguientes, que pueden encontrarse descritos en Pajares y Cruz (2007a): 1) estimador de máxima verosimilitud, modelado como una función de densidad de probabilidad

Gaussiana; 2) funciones de densidad de probabilidad gaussianas mixtas; 3) k -vecinos más próximos y 4) máquinas de vectores soporte. Estos clasificadores se entrenan de forma independiente para cada una de las clases identificadas. Posteriormente, para la clasificación, se combinan de forma que la decisión se toma en función de cuál de ellos proporciona la mejor opción. Aparentemente, este esquema parece sencillo, si bien la novedad más importante que introduce consiste en el hecho de que ante la posibilidad de que aparezca una muestra que no pertenece a ninguna de las clases utilizadas durante el entrenamiento, esta muestra se rechaza sin ser asignada a ninguna de las clases. El método se aplica a imágenes de satélite procedentes de dos sensores, Landsat TM (2009) y SAR (2009) (“Synthetic Aperture Radar”). En esta misma línea se sitúa el trabajo de Briem y col. (2002) para el mismo tipo de imágenes.

Fauvel y col. (2006) combinan los clasificadores mediante operadores de agregación fuzzy para clasificación de áreas urbanas a partir de imágenes del satélite IKONOS (SIC, 2009). Estos operadores son básicamente tres: combinación conjuntiva (mínimo), combinación disyuntiva (máximo) y un tercero que elige primeramente el mínimo y luego el máximo.

En Debeir y col. (2002) se propone un clasificador combinado, que utiliza la técnica conocida como “bagging” (Duda y col., 2001) con un método propio basado en la extracción de subconjuntos de características de los píxeles a clasificar. Los resultados obtenidos se comparan con el método k -NN, siendo k igual a cinco y también con el método basado en árboles de decisión C4.5. Se aplican a imágenes de satélite procedentes del sensor Landsat TM (2009).

En Kumar y col. (2002) se plantea la clasificación de imágenes proporcionadas por el sensor AVIRIS (2009), (ver sección 1.1). En dicho trabajo se utilizan 180 bandas espectrales, de forma que la dimensión del espacio de características para cada píxel se corresponde con dicho número de bandas. Debido a la alta dimensionalidad de los datos el número de clases es también elevado, produciéndose fuertes solapamientos entre las clases. La propuesta para la solución a esta problemática la proporciona el citado trabajo donde se establece una jerarquía en forma de árbol comenzando por el nodo raíz, que representa el problema original con todas las clases. El siguiente nivel de la jerarquía contiene una primera partición del conjunto de clases en subconjuntos disjuntos, obtenidos según ciertos criterios de separabilidad entre las clases. El proceso continúa

hasta que no se admiten más particiones. De este modo se consigue la separación de las clases, de forma que en cada nivel se estima una función de densidad de probabilidad para cada clase por máxima verosimilitud. Con dicha función y la probabilidad a priori se obtiene la probabilidad a posteriori mediante la regla de Bayes, que proporciona a su vez la probabilidad a priori para el siguiente nivel. La propia regla de Bayes, a través del producto de las probabilidades, constituye el método de combinación.

Petrakos y col. (2001) definen, en primer lugar, tres índices para medir los resultados de la clasificación, los cuales se obtienen comparando dos a dos los clasificadores, para ello calculan los siguientes índices: muestras clasificadas correctamente por ambos clasificadores, muestras clasificadas correctamente por el primer clasificador e incorrectamente por el segundo, muestras clasificadas incorrectamente por el primero y correctamente por el segundo y muestras clasificadas incorrectamente por ambos. Dichos índices se denominan: proporción de concordancia, proporción de concordancia específica y estadística de kappa (κ). El método se aplica a imágenes de satélite procedentes de dos sensores, Landsat TM (2009) y SAR (2009), para cada imagen y cada clase a identificar se estima una función de densidad de probabilidad por separado. Luego, dada una clase, ambas funciones se combinan bajo un esquema de promediado ponderado mediante un coeficiente que asigna un peso específico a cada una de las dos funciones. Los resultados se comparan dos a dos entre los diferentes métodos individuales y combinados que estudia.

En Jiménez y col. (1999) se propone la combinación en dos niveles de los datos hiperespectrales procedentes de tres sensores: óptico, térmico e infrarrojo. En una primera instancia, los datos de cada sensor se convierten en un espacio de características apropiado, donde son escalados a un rango de valores. En el segundo nivel, se lleva a cabo la decisión mediante técnicas clásicas de combinación de clasificadores tales como votador mayoritario, reglas del máximo, mínimo, media, todas ellas descritas en el capítulo tres, sección 3.5. El procedimiento combinado se aplica a imágenes de satélite en general.

Bruzzone y col. (1999) proponen una combinación de métodos para clasificar pares de imágenes obtenidas en diferentes instantes de tiempo y procedentes de diferentes sensores, por ejemplo un sensor óptico y otro SAR (2009). Ambas imágenes contienen la

misma región espacial. Los datos o características de los píxeles a clasificar procedentes de cada uno de los sensores constituyen las entradas para dos redes neuronales del tipo retro-propagación, las cuales, tras el correspondiente entrenamiento, proporcionan sendas salidas. Con estas salidas se estiman a su vez dos funciones de densidad de probabilidad utilizando el algoritmo conocido como Expectación Máxima (Duda y col., 2001). Previamente se determinan las probabilidades *a priori* teniendo en cuenta la frecuencia de aparición de las distintas clases en las imágenes analizadas. La función de densidad de probabilidad conjunta se obtiene considerando la propiedad de independencia entre ambas, resultando ser el producto de las dos funciones. Ante la llegada de una nueva muestra, se obtienen sendas probabilidades *a posteriori*. Las dos probabilidades se combinan mediante la regla de Bayes (Pajares y Cruz, 2007a) para tomar una decisión respecto de la clasificación para la muestra entrante.

En Wilkinson y col. (1995), Kanellopoulos y col. (1993) o Benediktsson y Kanellopoulos (1999) se propone la combinación de redes neuronales con métodos estadísticos basados en máxima verosimilitud bajo la hipótesis de que ambos métodos producen buenos resultados y por tanto, pueden complementarse. Inicialmente, los dos clasificadores se entrenan con el mismo conjunto de datos. Posteriormente, una segunda red neuronal se vuelve a entrenar sólo con aquellos datos en los que los dos clasificadores previos discrepan. Esta segunda red tiene la misión de discriminar píxeles ambiguos. Estos métodos se han probado en imágenes procedentes de diferentes sensores de satélites.

Los clasificadores descritos en el capítulo uno, sección 1.5, donde se exponen las aportaciones realizadas en el marco de esta investigación pertenecen a la categoría de esta sección, es decir, clasificadores combinados para segmentación de texturas en imágenes naturales de exterior.

2.4 Características de las imágenes en la clasificación de texturas naturales

Por característica se entiende un elemento significativo de la imagen, pudiendo ir desde una región formando parte de un objeto hasta un píxel, pasando por puntos significativos tales como bordes o incluso agrupaciones de píxeles siguiendo un cierto patrón. En cualquier caso, cada característica se describe por una serie de propiedades que la

identifican frente al resto. Las características y sus propiedades, constituyen las muestras para el proceso de clasificación, incluyendo el entrenamiento y la decisión. En la literatura se ha estudiado la conducta de diferentes características en la clasificación de las texturas (Valdovinos y col. 2005, Puig y García, 2006 o Hanmandlu y col. 2004). Existen dos enfoques: 1) métodos basados en las características espectrales de los píxeles (Puig y García, 2006; Rud y col. 2006; Hanmandlu y col. 2004; Yu y col. 2002; Kumar y col., 2001) y 2) métodos basados en regiones (Puig y García, 2006; Randen y Husøy, 1999; Wagner, 1999; Smith y Burns, 1997). Los métodos basados en los píxeles tratan de asignar el píxel a una clase determinada en función de sus propiedades específicas e individuales. Los métodos basados en las regiones tienen en cuenta las distribuciones espaciales de los píxeles dentro de la textura y por tanto su relación con otros píxeles en la región. En este último caso, cabe todavía distinguir entre texturas que presentan ciertas distribuciones regulares o patrones repetitivos, donde la unión de patrones define la textura global y aquellas otras que carecen de distribuciones de este tipo, tratándose por tanto, de texturas de naturaleza irregular.

En esta sección se analizan métodos basados en ambos enfoques. Así en la subsección 2.4.1 se revisan los métodos basados en los píxeles, bajo el punto de vista de su tratamiento individual. En la 2.4.2 se abordan los métodos que tienen en cuenta la distribución regular de las texturas, posponiendo el estudio de las texturas de naturaleza irregular para la sección 2.5. La intención es tratarlas por separado, dado que constituyen la base fundamental de este trabajo de investigación, puesto que estas texturas irregulares son las que aparecen en las imágenes naturales utilizadas en los experimentos llevados a cabo por los métodos propuestos en la investigación. Finalmente, en la subsección 2.4.3 se abordan las técnicas que utilizan tanto los píxeles de forma individual como formando parte de regiones.

2.4.1 Características espectrales a nivel de píxel

Como bien es sabido, los elementos básicos constituyentes de las imágenes son los píxeles, en este caso, considerados como características fundamentales de las propias imágenes, cuya propiedad más relevante es su naturaleza espectral. Existen técnicas de segmentación específicamente dedicadas a imágenes monocromas, es decir, imágenes de

grises. No obstante, lo más frecuente es la utilización del color como atributo (Koschan y Abidi, 2008; Pajares y Cruz, 2007a; Du y col., 2004; Cheng y col. 2001).

El tratamiento del color ha recibido una importante atención en las últimas dos décadas, en su calidad de propiedad descriptora de los píxeles en las texturas (Jiji y Ganesan, 2008; Rud y col. 2006; Luo y col., 1998; Comanicu y Meer, 1997; Pappas, 1992; Chang y col., 1994). Siguiendo esta línea, se han realizado diferentes estudios relacionados con los modelos de color, destacando los trabajos de Broek y Rikxoort (2004), Yu y col. (2002) o Kumar y col. (2001) donde se analizan entre otros, los modelos de color RGB (R, red; G, green y B, blue), HSI (H, hue, S, saturation y I, intensity), YIQ, YUV o LUV, todos ellos encaminados hacia la clasificación de texturas. En general, destacan como los más eficientes los modelos RGB y HSI. El modelo RGB es muy utilizado en técnicas basadas en gráficos por computador y específicamente en visión por computador. Posee una gran dependencia de las características del computador donde se visualiza o representa el gráfico o la imagen. Por otro lado, el modelo HSI está más bien relacionado con el concepto de percepción humana según se indica en Tkalcic (2003). En la literatura se han considerado otras representaciones de color tales como $L^*a^*b^*$ o $L^*u^*v^*$ debido también a su alto contenido perceptual (Yang y col., 2008).

En Trias-Sanz (2008) se hace un estudio sobre diferentes modelos de color para la segmentación de imágenes de satélite correspondientes a entornos rurales, llegando a la conclusión de que en algunos casos resulta más conveniente utilizar ciertos modelos de color transformados, además del modelo RGB, que también es empleado.

Un análisis comparativo, con un elevado nivel de profundidad, sobre distintas representaciones del color se puede encontrar en Cheng y col. (2001), incluyendo las características monocromas. En dicho trabajo se concluyen una serie de ventajas e inconvenientes que se reproducen en la tabla 2.1.

El análisis de diferentes trabajos, en segmentación de texturas en imágenes, lleva a la conclusión de que los modelos RGB y HSI compiten ambos por ser utilizados en dicha tarea. Aunque en algunos casos puntuales con el modelo RGB se pueden obtener peores resultados que con el HSI, bien es cierto que aquél, en general, se presenta como un modelo satisfactoriamente aceptable, incluso a pesar de que sus componentes poseen una cierta dependencia lineal entre ellas (Comanicu y Meer, 2002).

En algunos trabajos muy preliminares, como el de Robinson (1977), se llegó a la conclusión de que el modelo HSI era el más apropiado para la extracción de bordes en imágenes en color, si bien no es éste el problema que nos ocupa en relación a la clasificación de texturas abordado en este trabajo.

Tabla 2.1 Ventajas e inconvenientes de los diferentes modelos de color

Modelo de color	Ventajas	Inconvenientes
RGB	Conveniente para dispositivos de visualización gráfica en computadores	Componentes espectrales correladas
YIQ	Útil para monitores de TV del sistema Americano	Persiste la correlación espectral debido a la transformación lineal desde el modelo RGB
YUV	Útil para monitores de TV del sistema Europeo	Persiste la correlación espectral debido a la transformación lineal desde el modelo RGB
HSI	Basado en el sistema de percepción humano. Útil en algunos casos donde la iluminación varía debido a que la componente H es invariante a ciertos tipos de sombras. Es útil si la iluminación varía en una imagen o entre imágenes sucesivas.	Presenta fuertes singularidades cuando en las ecuaciones de transformación el denominador se aproxima a cero, lo que ocurre con bastante frecuencia. Es muy inestable con valores bajos de saturación.
Representaciones CIE	Se puede controlar la información del color y la intensidad independientemente.	Presenta fuertes singularidades debidas también a las ecuaciones de transformación desde el modelo RGB

Dentro del análisis del color existen otras aproximaciones relativas a la percepción visual, este es el caso de los trabajos propuestos por Vazquez y col. (2007) o Tkalcic (2003). En ellos, a partir del histograma se identifican pequeñas variaciones del mismo con el fin de suprimirlas, consiguiendo así un cierto suavizado, y con ello la homogenización de las regiones de texturas.

Montiel y col. (2005) proponen utilizar la información proporcionada por las distribuciones de los niveles espectrales a través del análisis de los correspondientes histogramas de las imágenes, que contienen las texturas.

Wangenheim y col. (2007), siguiendo la metodología basada en la división y fusión de regiones, aplican medidas de similitud para separar las clases que son diferentes y

fusionar las que son similares. Las medidas de similitud se establecen en relación a las componentes de color.

2.4.2 Características con patrones espaciales

Este tipo de características tienen la particularidad de presentar agrupaciones de píxeles con ciertas distribuciones regulares y espaciales de patrones. Existen diversas bases de imágenes con este tipo de texturas, tal es el caso del álbum de texturas de Brodatz (1966), disponible en numerosos sitios web actualizada o la base de datos VisTex por citar sólo algunas. En la Figura 2.1 se muestran varios ejemplos de tales texturas.



Figura 2.1 Diferentes ejemplos de texturas procedentes de las bases de imágenes Brodatz y VisTex

Para el análisis de dichas texturas se utilizan diversos operadores de vecindad y filtros cuya descripción se puede encontrar en Pajares y Cruz, (2007a), Gonzalez y Woods, (2007), Escalera (2001), Wagner (1999), Randen and Husoy (1999) o Smith y Burns (1997) entre otros. La idea consiste en situar una máscara alrededor de un determinado píxel, gracias a la cual se obtiene información sobre la distribución de los píxeles de la textura en la vecindad del píxel dado.

A continuación se ofrece una relación de diferentes métodos utilizados para la descripción de este tipo de texturas:

- Matrices de co-ocurrencia, con una serie de propiedades basadas en el histograma de las matrices, tales como: varianza, entropía o momentos de segundo orden entre otros (Chamorro-Martínez, 2007, Debeir y col. 2002).

- Filtros de Gabor (Wan, 2007; Palm y Lehmann, 2002; Dunn y Higgins, 1995; Porat y Zeevi, 1989).
- Máscaras de Laws (Gonzalez y Woods, 2007).
- Descomposición mediante la transformada de wavelets (Wang y Boesch, 2007, Pi y col. 2006; Permuter y col., 2006; Arivazhagan y Ganesan, 2003; Chang y Kuo, 1993; Unser, 1995).
- Descomposición en coeficientes de energía (Chen y col. 2005; Manjunath y Ma, 1996).
- Dominio de la frecuencia mediante análisis de Fourier (Lillo y col., 2007; Aiyer y col., 2005; Pyun y col., 2002, Clark y col., 1999).
- Transformada Discreta del Coseno (Li y col., 2006; Permuter y col., 2006).

En algunos casos se fusionan diversas estrategias, como ocurre en el método combinado descrito en Zhang y Wang (2004), que propone una estrategia multi-resolución mediante la descomposición de la imagen original en bandas de frecuencia a través de la transformada de wavelets, llegando a un cierto nivel de descomposición que es donde se utilizan descriptores basados en las matrices de co-ocurrencia (Pajares y Cruz, 2007). En Chen y col. (2005) también se utiliza una estrategia multi-resolución para la descomposición de la imagen en diferentes bandas de frecuencia.

En Trias-Sanz y col. (2008) se han analizado diferentes tipos de descriptores de textura, tales como filtros de Gabor, patrones locales binarios, estimadores de Fourier o intensidad media, incluyendo algunos otros de naturaleza estructural. Las imágenes utilizadas son de satélite con predominio de zonas rurales. En este trabajo se concluye que el comportamiento de dichos descriptores no resulta apropiado para este tipo de imágenes.

En la figura 2.2 se muestra una imagen representativa del conjunto de imágenes que se utilizan en el presente trabajo de investigación. Dicha imagen es cortesía del Servicio Territorial de Galicia en colaboración con la empresa Dimap (2009).

A la vista de dicha imagen, se deduce que las texturas presentes en la misma difieren sustancialmente del tipo de texturas mostradas en la imagen de la figura 2.1. Las diferencias más significativas se resumen a continuación:

1. En la figura 2.2 no aparecen patrones de textura repetitivos como en la figura 2.1, que permitan aplicar los descriptores que se usan para el tratamiento de esta última.
2. No existe una única textura, puesto que se distinguen claramente diversas zonas, tales como áreas boscosas, de praderas, zonas sin cultivar o construcciones asociadas a núcleos de población.
3. Existe una clara distinción de los colores asociados a las bandas espectrales.



Figura 2.2 Ejemplo de imagen aérea con diferentes texturas naturales

En cualquier caso, el tratamiento de la información espacial, según los patrones de textura vistos aquí, no debe confundirse con los métodos descritos en la sección 2.5, donde las distribuciones espaciales carecen de tales patrones, como se verá allí.

2.4.3 Combinación de características espectrales y espaciales

Las técnicas combinadas surgen bajo la idea de que la unión de esfuerzos resulta más productiva que los aportes individuales. En este caso, se trata de aprovechar la información conjunta proporcionada tanto por las características espectrales como la procedente de los patrones espaciales introducidos en la sección previa. Una herramienta

muy utilizada para llevar a cabo esta fusión de información es la transformada de wavelets.

En la línea expuesta previamente se sitúa el trabajo de Chen y col. (2005). En él se utilizan dos tipos de características. Unas describen la composición local del color a nivel de píxel y las otras las características espaciales de textura en la imagen de intensidad. Cada tipo de características se procesa por separado, para posteriormente proceder a su combinación. Las características de textura se obtienen descomponiendo la imagen en diferentes bandas de frecuencia, de forma que en cada nivel de la descomposición se utilizan aspectos relacionados con la frecuencia espacial.

En Wang y Boesch (2007) se combinan, tanto descriptores de color como de textura para la delimitación de masas forestales a partir de imágenes de satélite. Inicialmente la imagen se clasifica basándose exclusivamente en las componentes espectrales de color, donde cada píxel se etiqueta de acuerdo a la clase a la que pertenece, utilizando la técnica de división y fusión de regiones (Pajares y Cruz, 2007; Gonzalez y Woods, 2007; Koschan y Abidi, 2008). Tras la clasificación inicial, se determinan los bordes de las regiones en base a las propiedades derivadas de los coeficientes de textura obtenidos mediante la transformada de wavelets.

En Wan y col. (2007) se combina el color y la textura utilizando los descriptores del estándar MPEG-7 para la extracción del contenido visual de las imágenes con texturas naturales.

Trias-Sanz y col. (2008) combinan de manera jerárquica las propiedades espectrales con los descriptores de texturas. En su estudio analizan diversos modelos de color combinados a partes iguales con diversos descriptores de texturas. En primer lugar obtienen una segmentación inicial mediante el color y seguidamente realizan una homogeneización de las regiones basándose en la distribución espacial de los descriptores de textura.

Ozden y Polat (2007) combinan dos tipos de propiedades: *a*) componentes espectrales de cada píxel y *b*) valores obtenidos a partir de los descriptores de textura. Todo ello considerando la información espacial proporcionada por la vecindad de cada píxel. La información procedente de los descriptores de textura la obtienen a través de los

coeficientes de la transformada de wavelets (Pajares y Cruz, 2007a), considerando los coeficientes de detalle obtenidos mediante dicha transformada. Para incorporar la información espacial utilizan el procedimiento conocido como Mean-Shift (Comanicu y Meer, 2002).

En Chanussot y col. (2006) se propone una estrategia que utiliza en un primer nivel, la información espectral para clasificar imágenes urbanas obtenidas mediante el satélite IKONOS (SIC, 2009). En un segundo nivel utiliza la información espacial, que en este caso es información estructural, en el sentido de que aprovecha cierta información de este tipo para mejorar la clasificación. Exactamente, establece relaciones tales como que los tejados de las casas deben tener una forma geométrica determinada, las carreteras son estructuras alargadas, etc. De este modo, un píxel incorrectamente clasificado según alguna de las clases que identifican dichas formas se puede subsanar por la consideración de este tipo de información estructural.

2.5 Métodos de clasificación mediante información espacial

Aunque el tema de la incorporación de la información espacial para la clasificación de texturas ha sido tratado previamente, en esta sección se enfoca desde una perspectiva más general, abordando así las texturas espacialmente irregulares, es decir, sin patrones predefinidos. En esto se diferencian de las texturas estudiadas en las secciones 2.4.2 y 2.4.3 donde los descriptores de textura se basan precisamente en distribuciones espaciales regulares.

Las relaciones espaciales se pueden establecer de diversas maneras. De entre todas ellas destaca el enfoque basado en los denominados *Campos Aleatorios de Markov* (*Markov Random Fields*, MRFs) (Geman and Geman, 1984; Li, 2000). Además, las relaciones espaciales que se establecen en base a los MRFs son similares a las definidas en la propuesta que se hace en este trabajo de investigación en la sección 4.3. Por esta razón, en la sección 2.5.1 se realiza una revisión en profundidad de este enfoque. Las relaciones espaciales pueden establecerse también en función de ciertas propiedades estructurales, por ello en la sección 2.5.2 se consideran relaciones de esta naturaleza. Finalmente, en la sección 2.5.3 se revisan otros métodos, que establecen relaciones espaciales sin encuadrarse bajo una perspectiva concreta.

2.5.1 Información espacial mediante Campos Aleatorios de Markov

Básicamente, un MRF proporciona medidas de probabilidad sobre un dominio que posea relaciones espaciales. Este es el caso de las imágenes, ya que los objetos de la escena se proyectan en la imagen manteniendo sus relaciones espaciales. Esto es, los puntos vecinos en la escena 3D se proyectan también como píxeles vecinos en la imagen. El conjunto de posiciones donde se define un MRF se denomina rejilla. En las imágenes, la rejilla se genera a partir de los píxeles, que definen dichas localizaciones. Por consiguiente, bajo esta perspectiva, rejilla e imagen son entidades equiparables. Sobre cada localización espacial, definida por el correspondiente píxel de la imagen, se define una región de vecindad, que establece las relaciones entre las localizaciones que pertenecen a dicha región. A su vez, sobre la región de vecindad se define una función de energía local. La suma de todas las energías locales a lo largo de la rejilla, define una energía total. El objetivo de las técnicas basadas en los MRF consiste en minimizar la energía total. Cada posición de la rejilla se caracteriza por poseer una probabilidad, que puede obtenerse a partir de la energía.

Puesto que el caso que nos ocupa es un problema de clasificación de texturas, a continuación se proporciona un ejemplo ilustrativo de la aplicación de los MRFs en este contexto. Supongamos una región de vecindad sobre la rejilla de dimensión 3×3 centrada en la localización i definida sobre el píxel del mismo nombre. A dicha región pertenecen las ocho localizaciones vecinas, entre ellas k . Los píxeles asociados con dichas localizaciones han sido clasificados previamente de forma que se les ha asignado una etiqueta de clase, l_i y l_k pertenecientes al conjunto de etiquetas posibles. Esto es, disponiendo de c clases, identificadas como w_h con $h = 1, \dots, c$, entonces, si el píxel i pertenece a la clase w_h la etiqueta l_i es h , y del mismo modo, si el píxel k pertenece a la clase w_j la etiqueta l_k es j .



Figura 2.3 Relación de vecindad en un Campo Aleatorio de Markov

Se puede definir la función de energía total mediante la expresión siguiente,

$$E_L(l) = -\sum_L S(l_i, l_k) \text{ con } S(l_i, l_k) = \begin{cases} +1 & \text{si } l_i = l_k \\ -1 & \text{si } l_i \neq l_k \end{cases} \quad (2.1)$$

donde L se refiere a la región de vecindad alrededor de la localización i , siendo en este caso, $L = 3 \times 3$. A partir de la ecuación (2.1), se deduce que cuando ambas etiquetas tienen el mismo valor, $S(l_i, l_k)$ toma el valor $+1$. Si todas las etiquetas, en la vecindad de i , toman el mismo valor que l_i , entonces la energía resulta ser $E_L(l) = -8$. En este caso lo que se expresa es que todos los píxeles de la vecindad han sido clasificados como pertenecientes a la misma clase que el píxel central. Por el contrario, si todas las etiquetas de las localizaciones, alrededor de la localización central, difieren de ésta, entonces $S(l_i, l_k) = -1$ y $E_L(l) = +8$. Por tanto, en caso de compatibilidad en la clasificación de los píxeles en la vecindad, la energía es mínima, mientras que en caso de discrepancia total es máxima, siendo este uno de los planteamientos clave de los MRFs.

A partir de la energía local, E_L , se puede asignar a la localización i una cierta probabilidad de que la etiqueta l_i tome un valor concreto L . Por ejemplo, de acuerdo con la distribución de Gibbs se tendría la siguiente expresión (Haykin, 1994),

$$p(l_i = L) = Z^{-1} \exp \left\{ -\frac{1}{T} E_L(l) \right\} \quad (2.2)$$

Z es una constante de normalización y T es otra constante análoga a la temperatura en procesos del tipo de enfriamiento simulado, ambas pueden establecerse empíricamente. Bajo las suposiciones establecidas previamente, dada una localización i con etiqueta l_i , se deduce que a mayor compatibilidad entre i y sus vecinos, menor energía y por tanto mayor probabilidad asignada a la localización i . En caso de discrepancia entre i y sus vecinos, ocurre lo contrario, esto es, mayor energía y menor probabilidad.

Las diferencias entre los métodos que se citan a continuación, donde los MRFs constituyen su fundamento, estriban en el cómputo de esta probabilidad y naturalmente en cómo se combina esta probabilidad durante la clasificación según el método utilizado.

En Deng y Clausi (2004, 2005a,b) la energía consta de dos términos. Un término se obtiene a partir de las relaciones entre los píxeles por medio de sus características espectrales, de forma que la compatibilidad en la vecindad, se obtiene según las similitudes entre las características espectrales entre el píxel central y sus vecinos. El segundo término se obtiene a partir de las etiquetas de clase de una forma similar a la descrita previamente. Con ellos se calculan dos valores de probabilidad, que se combinan mediante el producto de ambos, obteniendo finalmente un único valor de probabilidad. El proceso de actualización de las probabilidades resultantes del producto se lleva a cabo mediante un procedimiento iterativo hasta conseguir el mínimo valor posible de energía.

El trabajo de Xia y col. (2006) sigue un enfoque similar al expresado previamente en los trabajos de Deng y Clausi (2004, 2005a,b), si bien abordan específicamente un problema importante que surge en los bordes de las regiones donde las texturas rompen su homogeneidad como consecuencia del tránsito de una región a otra. En este caso, en lugar de extraer primero las características de las regiones para proceder a su clasificación posterior, que suele ser una práctica general, se alterna sucesivamente entre obtención de características y clasificación. No obstante, el método también calcula dos términos de energía, a partir de los que obtiene las probabilidades en cada localización, que son combinadas mediante el producto. La actualización de las probabilidades se lleva a cabo mediante el conocido procedimiento de *Enfriamiento Simulado* (Duda y col, 2001). Un esquema similar al de Xia y col. (2006) es el propuesto por Borges y col. (2007).

En la línea de los trabajos anteriores se encuentran las investigaciones de Cariou y Chehdi (2008) o Allili y Ziou (2007), donde se plantean métodos basados en multi-resolución y por tanto en la descomposición de la imagen en varios niveles. En el primer nivel, la imagen se divide en regiones regularmente espaciadas con una dimensión de ventana fija, a partir de dichas regiones se estima: a) el número de clases de texturas; b) las características que describen las texturas. En el segundo nivel se obtiene la energía asociada a las localizaciones espaciales, bajo el paradigma de los MRFs, de una forma similar al procedimiento descrito en el párrafo previo. El proceso de actualización de las probabilidades se realiza mediante el método conocido en terminología inglesa como *Expectation Maximization* (EM) (Duda y col., 2001).

En Pyun y col. (2007) se propone una estrategia en la que se estiman las funciones de densidad de probabilidad suponiendo distribuciones Gaussianas, lo que constituye una práctica común en muchos métodos de clasificación. La estimación de dichas funciones se lleva a cabo utilizando un método de cuantización vectorial (LVQ) en lugar de utilizar el clásico EM, su justificación se fundamenta en una convergencia más rápida que la requerida por el EM. Como en el trabajo previo, la información espacial se introduce considerando la vecindad bajo el paradigma de los MRFs.

Huang (1999) y Campadelli y col. (1997) extraen la información espacial basándose en la similitud de etiquetas con las que han sido clasificados los píxeles vecinos en relación al central. Se aplica un proceso de minimización de la energía mediante la red neuronal de Hopfield. El cálculo de la energía se obtiene mediante un procedimiento idéntico al descrito en la ecuación (2.1).

Algunas otras investigaciones en el ámbito de los MRFs, con planteamientos similares a los anteriores, son los trabajos de Pyun y col. (2007), Yang y Jiang (2003), Melas y Wilson (2002), Li y col. (2000), Barrer y Rayner (2000), Sarkar y col. (2000), Andrey y Tarroux (1998), Krishnamachari y Chellapa (1997), Manjunath y Chellapa (1990).

2.5.2 Relaciones estructurales

El trabajo de Chen y col. (2005) incorpora la información espacial mediante la inyección de conocimiento procedente de la percepción humana. Se parte de los niveles elementales de color, a nivel de píxel (proceso local), para posteriormente utilizar descriptores de textura que incorporan información semántica. Un ejemplo de este tipo de información se refiere por ejemplo a situaciones tales como que las texturas correspondientes al cielo se situarán normalmente en la parte superior de la imagen y que en las proximidades de un lago habrá texturas procedentes de zonas de vegetación.

En al ámbito de las imágenes de satélite, la información espacial se ha incorporado también aplicando un conocimiento de índole estructural, lo que supone la aplicación de inteligencia de nivel superior. Los píxeles de una determinada estructura están rodeados de píxeles de esa misma estructura o de estructuras compatibles, por ejemplo en una zona urbana no es probable encontrar un píxel asociado con una zona de cultivo perteneciente

al ámbito rural. En esta línea se encuentran los trabajos de Barnsley y Barr (1996) o Gong y Howarth (1990).

2.5.3 Otros enfoques

En este apartado se referencian varios métodos, que sólo tienen en común el hecho de la incorporación de información espacial, lo cual se lleva a cabo bajo distintas perspectivas, que a su vez difieren de las expresadas anteriormente. Algunos de ellos se han utilizado en otros contextos diferentes al de las imágenes de texturas en entornos de exterior, sin embargo se incluyen aquí como ejemplos adicionales para la incorporación de la información espacial.

El método de Xie y col. (2004) se basa en sustituir el procedimiento de actualización iterativo EM por el clásico método conocido como Agrupamiento Borroso. Este último algoritmo se describe en el capítulo tres, sección 3.3.1 por ser uno de los métodos base utilizados en el presente trabajo de investigación. La aportación del trabajo de Xie y col. (2004) consiste en la definición de una nueva métrica, que incorpora tanto la información de similitud entre las características espectrales, como la información espacial. Esta última se obtiene mediante el cómputo de los grados de pertenencia de los píxeles vecinos, respecto de un píxel dado, a sus correspondientes clases. Cuanto mayor sea el grado de pertenencia de los vecinos, mayor es la contribución de éstos a la minimización de la función objetivo.

En la misma línea del anterior se encuentra la propuesta de Liew y col. (2003). En su trabajo, propusieron considerar los ocho píxeles vecinos al píxel central que se está tratando de clasificar; si los vecinos tienen valores similares y altos de pertenencia a la misma clase que el píxel central, el valor del grado de pertenencia de éste se modifica según la tendencia de sus vecinos. Podría considerarse como una especie de suavizado en función de la vecindad, por lo que su efecto sobre la modificación de los grados de pertenencia de píxeles que han sido incorrectamente clasificados tiene un efecto importante. Este método trata bastante bien las imágenes ruidosas de cara a una posible homogeneización de las texturas.

El trabajo propuesto por Krishnapuram y Freg (1992) incluye información de las coordenadas en el espacio 3D. Cuando dicha información está disponible, resulta de gran

ayuda para la clasificación, ya que los puntos situados en localizaciones espaciales contiguas tenderán a agruparse en localizaciones vecinas de la imagen bidimensional. Su principal inconveniente estriba en que se necesita tener disponible la información de las coordenadas espaciales con anterioridad al proceso de clasificación.

Kupinski y Gigger (1998), en el campo de las imágenes médicas, utilizaron una función isotrópica Gaussiana para interpretar el conocimiento a priori de cada forma de la región objetivo. La información espacial de la imagen se multiplica por esta función Gaussiana para eliminar valores de píxeles distantes, de tal forma que la imagen es posteriormente segmentada utilizando métodos de umbralización o técnicas de crecimiento de regiones (Pajares y Cruz, 2007). La información espacial incide en la mejora de las técnicas de umbralización o crecimiento de regiones. Desafortunadamente, a pesar de los buenos resultados mostrados, la generalización de este método no resulta fácil, particularmente en las imágenes de las texturas naturales utilizadas en nuestro trabajo donde los objetos no poseen formas determinadas dada la naturaleza de las texturas y por consiguiente la utilización de las semillas, como elementos clave en las técnicas de crecimiento de regiones, supone una dificultad añadida.

El uso de los contornos activos (*Active Contour Model*, ACM), es decir la delimitación de las regiones mediante contornos, ha sido otra de las técnicas utilizadas en la segmentación de las imágenes con texturas. El problema que entrañan es que necesitan un conocimiento previo de los objetos que se quieren identificar. Es evidente que en el caso de las texturas naturales esta suposición no es posible debido a que no existen objetos delimitados o con formas específicas y además las formas de las texturas difieren considerablemente de unas imágenes a otras. Por otro lado, requieren un elevado coste computacional, a pesar de las propuestas basadas en B-splines, que tratan de minimizar dicho coste (Precioso y col., 2005). Otras propuestas, que intentan reducir el coste computacional se encaminan hacia la segmentación de imágenes en varios niveles, de forma que en un nivel de resolución baja se obtienen los primeros resultados que guían el proceso en los niveles de más alta resolución (Kim y col., 2006).

2.6 Resumen del análisis bibliográfico

Recientemente, Hu y Damper (2008) han publicado un trabajo en el que tras el análisis de diversos métodos, se concluye que no existe ninguna estrategia combinada válida de

forma general. Así, dependiendo de la aplicación, una misma estrategia de combinación puede conducir a resultados diferentes. Esta consideración, junto con la revisión de métodos realizada en este capítulo, proporciona las pautas necesarias para abordar y estudiar nuevas estrategias de combinación de clasificadores, lo que constituye el fundamento del trabajo de investigación que se presenta en esta memoria.

Tras el análisis bibliográfico realizado previamente, se llega a las siguientes conclusiones, que darán soporte a los planteamientos propuestos en este trabajo:

1. Es necesario definir convenientemente las características a utilizar. Se han propuesto diferentes modelos de color, resultando la representación RGB una de las más apropiadas. Con la utilización de un espacio tridimensional, como es el caso del modelo RGB, se evitan algunos problemas relacionados con el hecho de usar espacios de alta dimensionalidad, tal y como se explica en Ho (2002). Una de las consecuencias derivadas de dicho trabajo estriba en el hecho de que a mayor dimensionalidad más difícil resulta la distribución de las muestras en clases.
2. Conviene utilizar técnicas combinadas frente al uso de métodos individuales. Esto está avalado por el mejor comportamiento obtenido con los métodos combinados frente a los métodos simples, tal y como se deriva de los estudios realizados en los trabajos referenciados en este capítulo, así como de las tendencias actuales.
3. Aunque los clasificadores supervisados puedan obtener mejores resultados que los no supervisados, debido a la intervención del propio supervisor, en aras de la automatización de los procesos, es preferible diseñar métodos de clasificación no supervisada por el mero hecho de tener que manejar abundantes datos.
4. Las imágenes utilizadas en este trabajo de investigación no poseen patrones de textura repetitivos, ni relaciones estructurales entre objetos de la escena. Por este motivo, la combinación de clasificadores puede aplicarse de forma local, esto es píxel a píxel.
5. No obstante, dadas las características espaciales de toda imagen, es posible utilizar este tipo de información global sobre la base de que dado un píxel, es probable que en su vecindad haya píxeles de su misma naturaleza, pertenecientes

a la misma región de textura o similares. En caso contrario delimitará zonas de diferentes texturas. Este es uno de los enfoques ampliamente utilizado en clasificación de texturas, como se ha puesto de manifiesto en los métodos basados en el paradigma de los MRFs.

En consecuencia, la propuesta realizada en este trabajo se centra en los siguientes aspectos:

1. Extracción de las características espectrales de los píxeles, en este caso, sus componentes R, G y B, generando así los vectores de características en el espacio tri-dimensional.
2. Diseño de un enfoque de clasificación no supervisado a partir de una estrategia de naturaleza supervisada.
3. Propuesta de dos estrategias de combinación a nivel local, esto es píxel a píxel, utilizando sendos enfoques para la combinación: *a)* la teoría de la decisión multi-criterio fuzzy y *b)* la integral fuzzy.
4. Propuesta de un enfoque de clasificación combinado de naturaleza global, bajo el cual se diseñan dos estrategias, la primera basada en los denominados Mapas Cognitivos Fuzzy y la segunda en un esquema de relajación probabilista. En ambas, la información espacial se incorpora teniendo en cuenta la vecindad de cada píxel. Se toman como referencia los modelos basados en los MRFs de donde se extraen las ideas fundamentales.
5. Valoración del comportamiento de las estrategias combinadas propuestas, frente a otros métodos de clasificación tanto individuales como combinados.

La estrategia no supervisada, junto con las propuestas combinadas locales y globales, se describen en el capítulo cuatro, mientras que la valoración de resultados se lleva a cabo en el capítulo cinco.

Capítulo 3

Clasificadores clásicos: individuales y combinados

3.1 Introducción

Este capítulo comienza con la descripción de seis clasificadores clásicos individuales de naturaleza supervisada, que constituyen la base para la construcción de otras estrategias de clasificación conocidas como combinadas. Tanto los métodos individuales como los combinados poseen una característica común, consistente en su funcionamiento en dos fases: *aprendizaje* y *decisión*. Esta es la razón por la que antes de describir los métodos de clasificación, se abordan en la sección 3.2 los aspectos relacionados con ambas fases.

En el párrafo anterior se alude al término clásico, que se refiere a la descripción del clasificador correspondiente según su versión original. El término individual expresa el hecho de utilizar un único clasificador, para distinguirlo de aquellas estrategias que emplean dos o más clasificadores, a las que nos referiremos como combinadas. Los seis clasificadores individuales, cuyos fundamentos se describen en la sección 3.3 son los siguientes (Pajares y Cruz, 2007a): *agrupamiento borroso*, estimación paramétrica mediante *máxima verosimilitud*, estimación no paramétrica mediante la *ventana de Parzen*, *mapas auto-organizativos* y finalmente sendas versiones del algoritmo de *cuantización vectorial*.

Como se verá en el capítulo cuatro, a pesar de la naturaleza *supervisada* implícita en los seis clasificadores anteriores, las propuestas combinadas que se formulan en el presente trabajo de investigación son de naturaleza *no supervisada*. Esto se consigue mediante el

diseño de una estrategia, que permite distribuir las muestras en clases, esto es, realizar una partición, junto con el correspondiente proceso de validación de dicha partición. Por tal motivo, es necesario en primer lugar, establecer el proceso para obtener la partición y en segundo lugar, fijar los criterios para su validación. La partición se obtiene mediante un procedimiento pseudo-aleatorio derivado a partir del clasificador de agrupamiento borroso y por consiguiente descrito en la sección 3.3. Los criterios de validación se definen en la sección 3.4.

En la sección 3.5 se describen varias estrategias clásicas combinadas. Por clásico, se entiende la combinación de los clasificadores individuales según la versión original del método utilizado, en contraposición a los métodos combinados propuestos en este trabajo de investigación, que se describen en el capítulo cuatro y que ya se han identificado en la sección 2.6 como dos estrategias de naturaleza local y otras dos de naturaleza global.

3.2 Fases de entrenamiento y decisión

Todo proceso de clasificación, en general y en particular la identificación de texturas en imágenes naturales, conlleva asociadas dos fases fundamentales: *entrenamiento* y *decisión*. Nos referiremos a la primera fase también como de aprendizaje, por identificarse ambos conceptos en la literatura especializada. Por la naturaleza de su procesamiento en el tiempo, a veces aparecen nombradas como procesos *off-line* y *on-line* respectivamente. Ello se debe al hecho de que la fase de entrenamiento se suele llevar a cabo durante periodos de inactividad del sistema, siendo en este momento cuando se estiman o aprenden los parámetros involucrados en el proceso. Por contra, la fase de decisión se realiza durante la plena operatividad del sistema, utilizando los parámetros aprendidos en la fase de entrenamiento.

La figura 3.1 muestra un esquema general de funcionamiento del sistema de aprendizaje/decisión particularizado para el caso de las imágenes de texturas naturales. Ambas fases constan tanto de procesos comunes como diferenciados. En efecto, los procesos de captura de imágenes, segmentación y codificación de la información son comunes, mientras que los procesos de Aprendizaje y Decisión son diferentes. Seguidamente se describen brevemente cada uno de ellos. Posteriormente, en cada método, bien sea individual o combinado, se establece la correspondiente diferenciación.

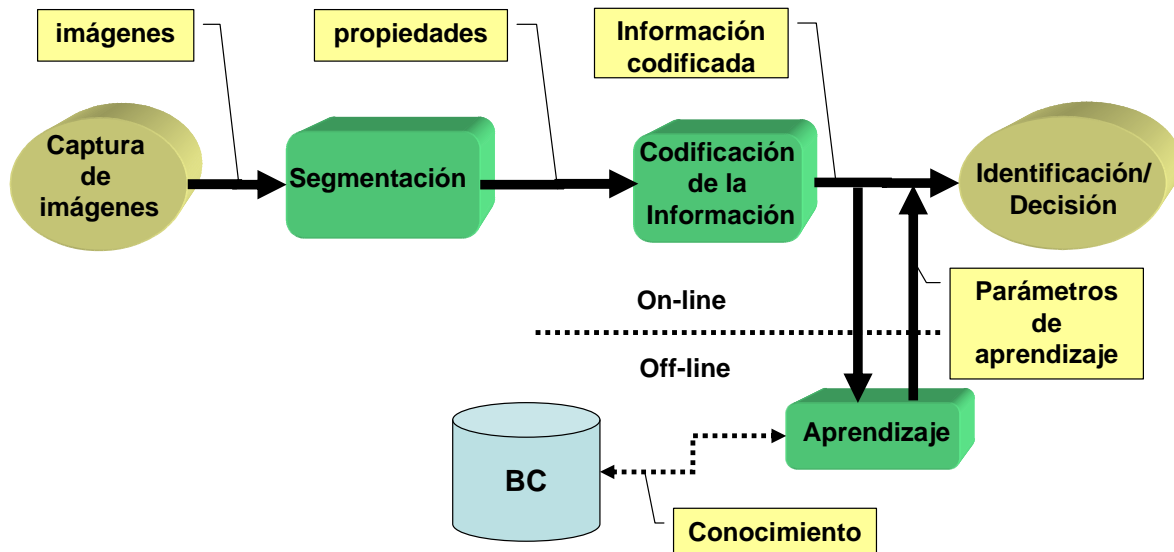


Figura 3.1 Esquema general de un proceso aprendizaje-decisión

En cualquier caso, como se ha indicado previamente, este esquema es válido tanto para los clasificadores de naturaleza individual como para los combinados.

3.2.1 Procesos comunes

Captura de imágenes: consiste en la obtención de las imágenes, bien tomadas de un almacén ya existente o bien directamente a partir de la escena mediante el correspondiente sensor.

Segmentación: se entiende por segmentación el procesado de la imagen o imágenes, consistente en la extracción de las estructuras o *características* subyacentes en ellas. Desde el punto de vista del tratamiento de imágenes, por característica se entiende algún elemento identificativo, sin que necesariamente tenga que coincidir con un objeto de la escena proyectado en la imagen. Así, una característica puede ser una región o un borde que pertenece a algún objeto sin necesidad de tener que llegar a identificar dicho objeto. Descendiendo un nivel más, una característica puede ser igualmente un píxel perteneciente a un borde, un punto de interés o sencillamente un píxel de la imagen independientemente de que pertenezca o no a alguna de las estructuras citadas. Los píxeles son las características utilizadas en este trabajo de investigación. De las características se deben extraer sus *propiedades* o *atributos* con el fin de proceder a su identificación. Siguiendo con lo expuesto, en el caso de una región puede ser su área, perímetro, intensidad media o cualquier otra propiedad descriptiva de la región. En nuestro caso, los atributos de los píxeles serán sus propiedades espectrales. En

consecuencia, el proceso de segmentación comprende tanto la extracción de características como sus propiedades.

Codificación de la información: esta fase incluye la estructuración de la información para ser posteriormente utilizada tanto por los métodos de aprendizaje como de clasificación. Cada característica extraída durante la fase anterior, se etiqueta convenientemente, procediendo a la cuantificación de sus propiedades en un rango de variabilidad apropiado. Las etiquetas son las *muestras*, representadas por vectores, cuyas componentes son las propiedades de la característica en cuestión. Para clarificar estos conceptos y ciñéndonos al rango de nuestra aplicación, tal y como se ha mencionado previamente, las características a considerar son los píxeles. Dado un píxel en la localización espacial (i, j) , si se etiqueta como k , se tiene $k \equiv (i, j)$, siendo su vector representante \mathbf{x}_k cuyas componentes son los valores espectrales de dicho píxel en el modelo de color RGB, es decir, $\mathbf{x}_k = \{x_{k_1}, x_{k_2}, x_{k_3}\} \equiv \{R, G, B\} \in \mathfrak{R}^3$ y por tanto, en este caso, el vector pertenece al espacio tridimensional \mathfrak{R}^3 . Las muestras así codificadas lo son tanto para el proceso de entrenamiento como para el de decisión, distinguiéndose entre muestras de entrenamiento y muestras para la clasificación, según la fase donde se procesen.

3.2.2 Procesos diferenciados

Aprendizaje/Entrenamiento: con las muestras disponibles convenientemente codificadas, se procede a la realización del correspondiente proceso de entrenamiento según el método seleccionado. Los procesos de entrenamiento se describen con detalle en este capítulo y el siguiente. Los parámetros resultantes del aprendizaje se almacenan en la Base de Conocimiento (BC), figura 3.1, para su posterior utilización durante la fase de decisión.

Identificación/Decisión: en esta etapa se procede a identificar una nueva característica o muestra, que todavía no ha sido clasificada como perteneciente a alguna de las clases de interés existentes. Para ello se recuperan los parámetros previamente aprendidos y almacenados en la BC; posteriormente y mediante la correspondiente función de decisión, inherente a cada método, se identifica la clase a la que pertenece. Este proceso también se denomina *reconocimiento* o *clasificación*. Incluso, en la literatura especializada, a veces aparece como segmentación, si bien en el caso que nos ocupa no utilizaremos esta

última denominación con el fin de diferenciarlo con nitidez del proceso de segmentación, expuesto previamente. Resulta a veces habitual el hecho de que las muestras una vez clasificadas, pueden incorporarse de nuevo al sistema, ahora ya como muestras de entrenamiento para proceder a un nuevo proceso de aprendizaje y por consiguiente llevar a cabo una actualización de los parámetros de aprendizaje asociados con cada método, que se almacenan de nuevo en la *BC*. Esto es lo que se conoce como aprendizaje incremental.

3.3 Métodos de clasificación clásicos individuales

En esta sección se describen los seis métodos de clasificación individuales mencionados en la introducción, que intervienen en el diseño de los clasificadores combinados propuestos en este trabajo de investigación y que volvemos a enumerar aquí por simplicidad: *agrupamiento borroso*, estimación paramétrica mediante *máxima verosimilitud*, estimación no paramétrica mediante la *ventana de Parzen*, *mapas auto-organizativos* y dos versiones de *cuantización vectorial*. Cada uno de ellos se describe tanto desde el punto de vista de la fase de entrenamiento como de la fase de decisión.

3.3.1 Agrupamiento borroso

3.3.1.1. Fase de entrenamiento

El objetivo de la técnica de agrupamiento conocida como *Agrupamiento borroso* o “*Fuzzy Clustering*” en terminología inglesa, consiste en realizar una partición de n elementos en c clases; cada uno de estos elementos es representado por el vector \mathbf{x} con q componentes, que representan las propiedades del mismo. Aunque el término fuzzy no está incluido en el diccionario de la lengua española, debido a su amplio uso, difusión y aceptación por parte de la comunidad científica, hemos optado por utilizarlo como tal, por lo que en este trabajo se consideran fuzzy y borroso términos sinónimos, utilizándose indistintamente, como ya se indicó anteriormente.

Con el fin de formalizar el método, sea el conjunto de muestras $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \in \mathfrak{R}^q$ un subconjunto del espacio real q -dimensional \mathfrak{R}^q . Cada $\mathbf{x}_k = \{x_{k_1}, x_{k_2}, \dots, x_{k_q}\} \in \mathfrak{R}^q$ se denomina vector de propiedades, siendo x_{k_j} la j -ésima propiedad o atributo de la muestra \mathbf{x}_k .

Como se ha mencionado previamente, y centrándonos en las imágenes que se utilizan en este trabajo de investigación, las características son los píxeles y sus propiedades las componentes espectrales en el modelo de color RGB , por tanto $q = 3$ y $\mathbf{x}_k = \{R, G, B\} \in \mathbb{R}^3$. Desde el punto de vista del clasificador, se habla de clases y muestras pertenecientes a las clases, en nuestro contexto de las imágenes naturales, las muestras son las características, es decir, los píxeles, mientras que las clases se identifican con las texturas que aparecen en las imágenes, cuyos detalles se explican en el capítulo cinco.

Este clasificador ha sido ampliamente utilizado en la literatura, cuya descripción se puede encontrar en Bezdek (1981), Zimmerman (1991), Duda y col., (2001) o Pajares y Cruz (2007a), entre otros. A continuación se proporciona una síntesis del mismo.

El objetivo de cualquier proceso de entrenamiento debe orientarse a conseguir que las muestras se distribuyan en clases, de tal forma que las pertenecientes a una misma clase sean tan similares entre sí como sea posible y a la vez diferenciarse de las muestras de otras clases, también tanto como sea posible. Por este motivo es necesario definir alguna medida de similitud, que en el caso que nos ocupa serán funciones de distancia, de forma que la distancia entre dos muestras \mathbf{x}_k y \mathbf{x}_l determine el grado de similitud entre ellas.

La distribución de las muestras del conjunto X en clases se conoce como *partición*. Si se pretende realizar una partición del conjunto X en c clases, se tendrán $w_i \{i=1, \dots, c\}$ subconjuntos de X , donde cada subconjunto constituirá una clase. Una partición puede contemplarse desde las dos perspectivas siguientes: *fuzzy* y *no fuzzy*. Una partición no fuzzy se conoce en terminología inglesa como “crisp”. Desde el punto de vista fuzzy, para cada clase w_i se puede definir un conjunto borroso μ_i sobre el universo X , $\mu_i : X \rightarrow [0,1]$ que asigna a cada muestra \mathbf{x}_k lo que se conoce como *grado de pertenencia* de dicha muestra al subconjunto w_i , denominado *clase* a partir de este momento. El grado de pertenencia de \mathbf{x}_k a la clase w_i se expresa como $\mu_i(\mathbf{x}_k) \equiv \mu_{ik}$. En el caso de conjuntos no borrosos, denominados “crisp” en la literatura inglesa, una muestra \mathbf{x}_k se dice que pertenece a la clase w_i dada y no pertenece al resto de clases. Esto se expresa mediante los valores discretos $\{0,1\}$ de la siguiente forma $\mu_{ik} = 1$ para

indicar que pertenece y $\mu_{ik} = 0$ para expresar que no pertenece. Por el contrario, en el caso de conjuntos borrosos se dice que una muestra puede pertenecer a diferentes clases y así se habla por ejemplo de que \mathbf{x}_k pertenece a una clase w_i con grado de pertenencia μ_{ik} y a la vez a la clase w_j con grado de pertenencia μ_{jk} . Como ejemplo, supongamos que se tienen tres clases w_i , w_j y w_h , en este caso podríamos decir que la muestra \mathbf{x}_k pertenece a dichas clases con los siguientes grados de pertenencia $\mu_{ik} = 0.4$, $\mu_{jk} = 0.5$ y $\mu_{hk} = 0.1$, donde estos valores pertenecen al intervalo continuo $[0,1]$.

Dado el conjunto de muestras $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ y el conjunto V_{cn} de todas las matrices reales de dimensión $c \times n$, con $2 \leq c < n$, se puede obtener una matriz de grados de pertenencia representando una partición de las muestras en clases de la siguiente manera $U = \{\mu_{ik}\} \in V_{cn}$. Tanto en el supuesto “crisp” como en el fuzzy se deben cumplir las siguientes condiciones:

$$\begin{aligned}
 a) \quad & \mu_{ik} \in \{0,1\} \text{ crisp} \quad \text{o} \quad \mu_{ik} \in [0,1] \text{ fuzzy} \quad 1 \leq i \leq c; \quad 1 \leq k \leq n \\
 b) \quad & \sum_{i=1}^c \mu_{ik} = 1 \quad 1 \leq k \leq n \\
 c) \quad & 0 < \sum_{k=1}^n \mu_{ik} < n \quad 1 \leq i \leq c
 \end{aligned} \tag{3.1}$$

Para ilustrar los conceptos anteriores, sea $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, entonces podemos construir las siguientes matrices suponiendo que $c = 2$, es decir se dispone de dos clases. En ambas matrices las filas se asocian con las clases, así si nos fijamos en U_1 diríamos que \mathbf{x}_1 pertenece a la clase w_1 con grado de pertenencia uno y no pertenece a la clase w_2 al tener un grado de pertenencia cero. Por el contrario, si consideramos U_2 , ahora \mathbf{x}_1 pertenece a w_1 con grado de pertenencia 0.9 pero simultáneamente, también pertenece a w_2 con grado de pertenencia 0.1.

“crisp”

$$U_1 = \begin{matrix} & \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} \\ \begin{matrix} w_1 \\ w_2 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \end{matrix}$$

“fuzzy”

$$U_2 = \begin{matrix} & \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} \\ \begin{matrix} w_1 \\ w_2 \end{matrix} & \begin{bmatrix} 0.9 & 0.4 & 0.2 \\ 0.1 & 0.6 & 0.8 \end{bmatrix} \end{matrix}$$

Una clase w_j se representa por su centro $\mathbf{v}_j = \{v_{j_1}, v_{j_2}, \dots, v_{j_q}\} \in \mathbb{R}^q$ teniendo en cuenta que $j = 1, \dots, c$. Intuitivamente, este centro sería el representante de las muestras en la clase, las cuales tienden a agruparse a su alrededor. Como se ha mencionado previamente en el caso de las imágenes utilizadas en este trabajo, q es igual a tres.

El proceso de entrenamiento o aprendizaje recibe las muestras del conjunto X , siendo conocido el número de clases, c . Las muestras se encuentran distribuidas en esas c clases. Dado el conocimiento de c , el proceso se engloba dentro de los métodos de clasificación de naturaleza *supervisada*. El objetivo final de este proceso consiste en estimar los centros de las clases así como los grados de pertenencia de cada muestra a cada clase. Con tal propósito se define la función objetivo J , que debe ser minimizada asumiendo las restricciones definidas en la ecuación (3.1) (Bezdek, 1981; Duda y col., 2001; Zimmerman, 1991),

$$J(U; \mathbf{v}) = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m d_{ij}^2 \quad (3.2)$$

donde $\mathbf{v} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c\}$, $\mathbf{v}_j \in \mathbb{R}^q$ representa el centro de la clase j -ésima; $d_{ij}^2 \equiv d^2(\mathbf{x}_i, \mathbf{v}_j)$ expresa una medida de distancia al cuadrado entre la muestra \mathbf{x}_i y el centro de la clase w_j dado por \mathbf{v}_j ; m se conoce como peso exponencial (Bezdek, 1981).

Para minimizar la función objetivo, los centros de las clases y los grados de pertenencia conviene elegirlos de modo que los valores altos en los grados de pertenencia se obtengan para aquellas muestras que estén cerca, en términos de distancia, de los correspondientes centros de las clases. Por otro lado, cuanto mayor sea el valor de m , menor será la contribución de las muestras con valores bajos de los grados de pertenencia a la función objetivo. Dichas muestras tienden a ser ignorados en la estimación de los centros de las clases y grados de pertenencia (Zimmerman, 1991). El parámetro m se ajusta durante la fase experimental mediante el procedimiento de validación cruzada (“cross-validation”) descrito en el capítulo cinco, sección 5.3 (Duda y col., 2001).

El algoritmo del clasificador borroso se fundamenta en un proceso iterativo, de forma que en cada iteración t , se modifican los centros de las clases en función de los grados de pertenencia que también se actualizan, según las ecuaciones (3.3) y (3.4),

$$\mu_{ij}(t+1) = \frac{1}{\sum_{r=1}^c (d_{ij}(t)/d_{ir}(t))^{2/(m-1)}} \quad (3.3)$$

$$\mathbf{v}_j(t+1) = \frac{\sum_{i=1}^n \mu_{ij}^m(t) \mathbf{x}_i}{\sum_{i=1}^n \mu_{ij}^m(t)} \quad (3.4)$$

El proceso se detiene en aquella iteración en la que se verifica que $\|\mu_{ij}(t+1) - \mu_{ij}(t)\| < \varepsilon \quad \forall ij$, lo que se conoce como criterio de convergencia. No obstante, en la práctica se suele establecer un criterio adicional fijando un número máximo de iteraciones, t_{max} , para que el proceso se detenga en caso de no conseguir la convergencia.

Los centros de las clases constituyen los parámetros de aprendizaje, que se almacenan en la *BC*, figura 3.1.

El proceso iterativo descrito previamente requiere, o bien comenzar con unos grados de pertenencia dados y a partir de ellos calcular los centros o a la inversa, esto es, comenzar con unos centros dados para posteriormente continuar calculando los grados de pertenencia. En este trabajo se ha optado por la segunda alternativa. Con tal propósito, se ha elegido el método pseudo-aleatorio propuesto por Balasko y col. (2006), gracias al cual se inicializan los centros calculando las medias estadísticas de las clases, obviamente suponiendo conocido el número de clases. De este modo, el proceso de convergencia comienza con unos valores de los centros próximos a los finales. La inicialización de los centros se lleva a cabo según los dos pasos siguientes:

1. Realizar una transformación lineal $Y = f(X)$ de las muestras de entrenamiento X , de tal forma que sus valores queden restringidos al rango $[0,1]$.
2. Inicializar el vector de centros $\mathbf{v} = 2D\bar{\mathbf{M}} \circ \mathbf{R} + D\bar{\mathbf{m}}$, donde $\bar{\mathbf{m}}$ es el vector media para la transformada de las muestras de entrenamiento con los valores de Y y $\bar{\mathbf{M}} = \max(\text{abs}(Y - \bar{\mathbf{m}}))$, ambos del tamaño $1 \times d$; $D = [1 \quad \dots \quad 1]^T$ con tamaño $c \times 1$; \mathbf{R} es una matriz de número aleatorios con valores en el intervalo $[0,1]$ y tamaño $c \times q$; la operación \circ define la multiplicación elemento por elemento de las matrices.

3.3.1.2. Fase de decisión

La decisión constituye el proceso por el cual una nueva muestra \mathbf{x}_s cuya pertenencia a las clases resulta desconocida hasta el momento, debe ser identificada como perteneciente a una de las clases w_j disponibles. Con tal propósito, se recuperan de la BC, figura 3.1, los centros de las clases \mathbf{v}_j ($j = 1, 2, \dots, c$) a partir de los cuales se obtienen los grados de pertenencia de \mathbf{x}_s a cada una de las clases según la siguiente expresión,

$$\mu_{sj} = \frac{1}{\sum_{r=1}^c (d_{sj}/d_{sr})^{2/(m-1)}} \quad (3.5)$$

donde d_{sj} define, como antes, una distancia entre la muestra \mathbf{x}_s y el centro de la clase \mathbf{v}_j , $d_{sj} \equiv d(\mathbf{x}_s, \mathbf{v}_j)$; en el caso que nos ocupa se ha elegido la distancia Euclídea. Los grados de pertenencia dados por la ecuación (3.5) constituyen las salidas de este clasificador.

De esta forma se obtiene la siguiente secuencia de grados de pertenencia para \mathbf{x}_s ,

$$\mu_{s1} \equiv \mu(\mathbf{x}_s, w_1), \mu_{s2} \equiv \mu(\mathbf{x}_s, w_2), \dots, \mu_{sc} \equiv \mu(\mathbf{x}_s, w_c) \quad (3.6)$$

Finalmente, \mathbf{x}_s se clasifica como perteneciente a la clase w_j que proporcione el máximo grado de pertenencia según la siguiente regla de decisión,

$$\mathbf{x}_s \in w_j / \mu_{sj} > \mu_{sk} \quad \forall k \neq j \text{ con } k, j = 1, 2, \dots, c \quad (3.7)$$

3.3.2. Estimación Paramétrica: máxima verosimilitud

3.3.2.1. Fase de entrenamiento

Este método tradicionalmente se ha venido encuadrando dentro de las técnicas de clasificación *no supervisada* (Escudero, 1977). Dada una muestra de entrenamiento genérica $\mathbf{x} \in \mathfrak{R}^q$, el objetivo consiste en calcular las probabilidades de pertenencia a cada una de las clases w_j , esto es $P(w_j | \mathbf{x})$.

Esta técnica parte de la base de que se conoce la función de densidad de probabilidad condicional para cada clase, resultando desconocidos los parámetros o estadísticos involucrados en dicha función. Una práctica muy generalizada, adoptada en este trabajo,

consiste en asumir que la forma de estas funciones sigue la ley de distribución Gaussiana o Normal, según la siguiente expresión,

$$p(\mathbf{x} | \mathbf{m}_j, C_j) = \frac{1}{(2\pi)^{q/2} |C_j|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{m}_j)^T C_j^{-1} (\mathbf{x} - \mathbf{m}_j) \right\} \quad (3.8)$$

donde \mathbf{m}_j y C_j son respectivamente la media y la matriz de covarianza de la clase w_j , o dicho de otra forma, los estadísticos o parámetros desconocidos que deben ser estimados; T denota la matriz transpuesta y q expresa la dimensionalidad de los datos según $\mathbf{x} \in \mathbb{R}^q$.

La hipótesis que, específicamente, asumen las técnicas de clasificación no supervisadas son las siguientes:

1. Existen c clases en el problema.
2. La muestra \mathbf{x} proviene de estas c clases, aunque se desconoce la clase específica a la que pertenece.
3. La probabilidad a priori de que la muestra pertenezca a la clase w_j , $P(w_j)$ resulta en principio desconocida.
4. La función de densidad asociada a cada clase posee una forma conocida, siendo desconocidos los parámetros de dicha función.

Con este planteamiento resulta factible aplicar la regla de Bayes para obtener la probabilidad condicional de que \mathbf{x} pertenezca a la clase w_j , mediante la siguiente expresión (Huang y col., 2002),

$$P(w_j | \mathbf{x}) = \frac{p(\mathbf{x} | \mathbf{m}_j, C_j) P(w_j)}{\sum_{j=1}^c p(\mathbf{x} | \mathbf{m}_j, C_j)} \quad (3.9)$$

Conocidas las formas de las funciones de densidad de probabilidad, el método de máxima verosimilitud tiene por objetivo estimar los mejores parámetros para dichas funciones.

En términos de probabilidad, si $p(\mathbf{x} | \mathbf{m}_j, C_j)$ es la probabilidad de que la muestra representada por su vector \mathbf{x} tome en la clase w_j dicho valor, entonces, teniendo en cuenta

los estadísticos de todas las clases, la probabilidad de la muestra \mathbf{x} será la suma de todas las probabilidades para las c clases, que viene dada por la siguiente expresión,

$$P(\mathbf{x} | \mathbf{m}_j, C_j) = \sum_{j=1}^c p(\mathbf{x} | \mathbf{m}_j, C_j) P(w_j) \quad (3.10)$$

donde $P(\mathbf{x} | \mathbf{m}_j, C_j)$ se conoce como probabilidad total (“mixture density”), las probabilidades condicionales $p(\mathbf{x} | \mathbf{m}_j, C_j)$ son los componentes de la probabilidad total (“component densities”) y las probabilidades a priori son los parámetros de la mezcla (“mixing parameters”).

Considerando la probabilidad total, la probabilidad de que se extraiga la muestra \mathbf{x} del colectivo viene dada por la probabilidad conjunta de cada muestra \mathbf{x}_i , según la expresión,

$$P(\mathbf{x} | \mathbf{m}_j, C_j) = \prod_{i=1}^n p(\mathbf{x}_i | \mathbf{m}_j, C_j) \quad (3.11)$$

siendo n el número total de muestras disponibles en el conjunto $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \in \mathfrak{R}^q$.

La estimación más acertada para los parámetros \mathbf{m}_j y C_j , según el método de máxima verosimilitud, será aquella que maximice la función dada en (3.11). Desde un punto de vista analítico es preferible trabajar con el logaritmo neperiano de dicha función, ya que al ser ésta una función monótona creciente, el vector que maximiza el logaritmo también maximiza la función de (3.11). Por tanto, las mejores estimaciones para \mathbf{m}_j y C_j se obtienen calculando las primeras derivadas del logaritmo de la función dada en (3.11) e igualándolas a cero, como condición de máximo.

En la expresión (3.11) se supone que las muestras en una clase w_h no proporcionan información sobre la estimación de los parámetros para la clase w_j , lo cual significa que los parámetros para las diferentes clases son funcionalmente independientes. Esto se traduce en que es posible trabajar con cada clase separadamente y en consecuencia el número de muestras total n en (3.11) puede sustituirse por n_j , que representa el número de muestras pertenecientes a la clase w_j , que intervienen en el cómputo de la función de densidad de probabilidad asociada a dicha clase. Por tanto, de aquí en adelante, y bajo la

suposición de que se tienen c problemas, el número de muestras que intervienen en el cómputo de los parámetros de la clase w_j será n_j (Duda y col., 2001).

Consideremos el supuesto univariable con $\mathbf{m}_j = m_j$ y $C_j = \sigma_j^2$ y $\mathbf{x}_i = x_i$, en cuyo caso,

$$\ln P(\mathbf{x} | m_j, C_j) = \sum_{i=1}^{n_j} \ln p(x_i | m_j, C_j) \quad (3.12)$$

teniendo en cuenta la ecuación (3.8) para el caso unidimensional, $q = 1$, resulta,

$$\ln p(x_i | m_j, C_j) = -\frac{1}{2} \ln 2\pi C_j - \frac{1}{2C_j} (x_i - m_j)^2 \quad (3.13)$$

Derivando la expresión (3.13) con respecto a cada uno de los parámetros que se están estimando, es decir m_j y C_j se tiene,

$$\frac{\partial}{\partial m_j} \ln p(x_i | m_j, C_j) = \frac{1}{C_j} (x_i - m_j) \quad (3.14)$$

$$\frac{\partial}{\partial C_j} \ln p(x_i | m_j, C_j) = -\frac{1}{2C_j} + \frac{(x_i - m_j)^2}{2C_j^2} \quad (3.15)$$

Considerando las n_j muestras de entrenamiento, de acuerdo con la discusión previa y según la ecuación (3.12), se obtiene,

$$\frac{1}{n} \sum_{i=1}^{n_j} \frac{1}{\hat{C}_j} (x_i - \hat{m}_j) = 0 \quad -\frac{1}{n_j} \sum_{i=1}^{n_j} \frac{1}{2\hat{C}_j} + \frac{1}{n_j} \sum_{i=1}^{n_j} \frac{(x_i - \hat{m}_j)^2}{2\hat{C}_j^2} = 0 \quad (3.16)$$

donde \hat{m}_j y \hat{C}_j son las estimas de máxima verosimilitud para m_j y C_j , respectivamente.

Resolviendo, obtenemos finalmente las estimaciones para los parámetros desconocidos de acuerdo con la siguiente expresión,

$$\hat{m} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i \quad \hat{\sigma}^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (x_i - m)^2 \quad (3.17)$$

Aunque el análisis del caso multivariable es básicamente similar, se requiere mucha más manipulación para su obtención. El resultado, muy bien conocido en estadística, es que las estimaciones de máxima verosimilitud para \mathbf{m}_j y C_j vienen dadas por,

$$\mathbf{m}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbf{x}_i \quad C_j = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^T \quad (3.18)$$

La ecuación (3.18) expresa que la estima de máxima verosimilitud para el vector media es la media aritmética simple, mientras que la estima de máxima verosimilitud para la matriz de covarianza es la media aritmética de las n_j matrices $(\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^T$.

A partir de este momento quedan totalmente especificadas las funciones de densidad de probabilidad definidas en la ecuación (3.8), ya que los parámetros, que allí eran desconocidos, acaban de ser estimados. Estos parámetros \mathbf{m}_j y C_j ($j = 1, 2, \dots, c$) se almacenan finalmente en la BC de acuerdo con el esquema de la figura 3.1.

3.3.2.2. Fase de decisión

Dada una nueva muestra \mathbf{x}_s , el objetivo consiste en clasificarla como perteneciente a una de las clases disponibles. Para ello se calculan las probabilidades condicionales de que \mathbf{x}_s pertenezca a cada una de las clases w_j de acuerdo con la expresión (3.9) según la siguiente regla de decisión (Duda y col., 2001),

$$\mathbf{x}_s \in w_j \quad \text{si} \quad P(w_j / \mathbf{x}_s) > P(w_k / \mathbf{x}_s) \quad \forall k \neq j \quad (3.19)$$

No obstante, en esta regla el denominador de la ecuación (3.9) no intervine en la decisión por ser idéntico en ambos miembros de la desigualdad; por consiguiente, la regla de decisión puede reescribirse como sigue,

$$\mathbf{x}_s \in w_j \quad \text{si} \quad p(\mathbf{x}_s / \mathbf{m}_j, C_j) P(w_j) > p(\mathbf{x}_s / \mathbf{m}_k, C_k) P(w_k) \quad \forall k \neq j \quad (3.20)$$

La única cuestión pendiente estriba en el cómputo de las probabilidades a priori. Cuando éstas no se conocen, normalmente la forma de proceder consiste en fijar todas las probabilidades a priori al mismo valor, por ejemplo a $1/2$, en cuyo caso, dejarían de participar en la decisión final. Esta es la asunción hecha en este trabajo, por lo que la expresión (3.20) se transforma en la (3.21),

$$\mathbf{x}_s \in w_j \text{ si } p_b(\mathbf{x}_s / w_j) > p_b(\mathbf{x}_s / w_k) \quad \forall k \neq j \quad (3.21)$$

No obstante, en algunas estrategias combinadas donde participa este clasificador, es necesario normalizar los valores de probabilidad al rango $[0, +1]$ para competir en igualdad de condiciones con el resto de clasificadores que intervienen en la combinación. Por este motivo, en sus actuaciones combinadas, la regla de decisión es la proporcionada por la ecuación (3.19), ya que gracias al efecto del denominador del segundo miembro de la ecuación (3.9) el resultado final está normalizado a dicho rango.

Por consiguiente y según lo anterior, en este trabajo de investigación, las salidas del clasificador son las probabilidades proporcionadas a través de la ecuación (3.9), si bien renombradas con el subíndice b , esto es $P_b(w_j / \mathbf{x})$ para diferenciarlas posteriormente de las probabilidades que se obtienen a través del estimador no paramétrico basado en la ventana de Parzen.

3.3.3. Estimación no paramétrica: ventana de Parzen

3.3.3.1. Fase de entrenamiento

En este proceso, al igual que en el caso paramétrico de máxima verosimilitud, el objetivo sigue siendo el cálculo de las probabilidades de pertenencia de la muestra \mathbf{x} a cada una de las clases w_j , esto es $P(w_j | \mathbf{x})$.

Por tanto, el problema se plantea desde el mismo punto de vista, formulando las mismas tres primeras hipótesis y sustituyendo la cuarta por una nueva que ahora se concreta como: “la forma de la función de densidad de probabilidad asociada a cada clase no se conoce”. Esto significa que en este caso no existen parámetros a estimar, siendo por tanto la propia función de densidad de probabilidad el objetivo de la estimación (Parzen, 1962, Duda y col. 2001). La estimación se lleva a cabo considerando la influencia de cada una de las muestras en la función de densidad de probabilidad a través de una función potencial, que determina el grado de influencia en la participación de cada muestra, tal y como se expresa en la ecuación (3.22) con n_j el número de muestras en la clase w_j ,

$$p_p(\mathbf{x} / w_j) = \frac{1}{n_j} \sum_{k=1}^{n_j} K_h(\mathbf{x} - \mathbf{x}_k) \quad (3.22)$$

En la ecuación anterior, $K_h(z) = K(z/h)/V_h$, define la función potencial a través de la denominada ventana o función del núcleo $K(\cdot)$, cumpliendo $\int_{-\infty}^{+\infty} K(k)dk = 1$. Esta propiedad establece que la función de densidad de probabilidad cumple que el área bajo ella debe ser igual a la unidad; h se conoce como parámetro de suavizado o ancho de la ventana, que determinará el intervalo de influencia del núcleo, con $h > 0$ (Fukunaga y Hummels, 1987). El subíndice p hace referencia a la probabilidad de Parzen, para distinguirla de la probabilidad basada en la estimación de máxima verosimilitud definida previamente e identificada con el subíndice b .

Típicamente, la función potencial se elige de forma que sea una función monótona decreciente, tal que disminuya con la distancia de su argumento a la muestra \mathbf{x}_j . Si llamamos a esta función decreciente g , entonces $K(\cdot)$ se puede expresar a través de g de forma que: $K(\mathbf{x} - \mathbf{x}_k) = g(d(\mathbf{x}, \mathbf{x}_k))$, donde $d(\mathbf{x}, \mathbf{x}_j)$ es una distancia métrica. Las propiedades estadísticas para este estimador se pueden encontrar en Duda y col. (2001).

La ecuación (3.22) expresa que la contribución de cada muestra de entrenamiento \mathbf{x}_k , a la estimación de la función de densidad en el punto \mathbf{x} está determinada por el ancho y la forma de la función del núcleo. Por ejemplo, si la ventana tiene la forma Gaussiana, las muestras que caen cerca de \mathbf{x} contribuyen más a la estimación de la función de densidad que aquellas otras que se sitúen lejos de \mathbf{x} . Tales ecuaciones son interpretadas como un proceso de centrado del núcleo $K(\cdot)$ sobre cada muestra de entrenamiento, posteriormente se suman las n_j funciones solapadas en el espacio alrededor de \mathbf{x} , de ahí el sumatorio que aparece en la ecuación (3.22).

A la hora de estimar la función de densidad de probabilidad, la forma del núcleo no es tan importante como su ancho. Se han propuesto diferentes tipos de núcleos (Duda y col., 2001). En este trabajo se ha elegido el núcleo Gaussiano, por las dos razones que se exponen a continuación. La primera es porque con este núcleo obtenemos toda la información sobre el grado de correlación entre los valores de las propiedades de las muestras de entrenamiento a través de la matriz de covarianza C_j , para la clase w_j . La segunda razón viene dada por sus excelentes propiedades analíticas. En consecuencia, la función de densidad finalmente estimada resulta ser la proporcionada por la ecuación (3.23).

$$p_p(\mathbf{x} / w_j) = \frac{1}{n_j} \sum_{k=1}^{n_j} \left\{ \frac{\exp\{-D(\mathbf{x}, \mathbf{x}_k, h_j)\}}{(2\pi)^{q/2} h_j^{n_j} |C_j|^{1/2}} \right\} \quad (3.23)$$

donde $D(\cdot) = (\mathbf{x} - \mathbf{x}_k)^T C_j^{-1} (\mathbf{x} - \mathbf{x}_k) / 2h_j^2$; q representa la dimensión de las muestras en el espacio considerado; T indica la operación de transposición de vectores. Cada función de núcleo considera que las muestras se sitúan alrededor del patrón teórico \mathbf{x}_k donde el vector promedio para cada núcleo es el propio patrón. El parámetro de suavizado h_j se expresa generalmente en función del número de patrones en la clase, esto es n_j ,

$$h_j = h n_j^{-\frac{r}{q}} \quad \text{para} \quad 0 \leq r \leq 0.5 \quad (3.24)$$

La elección del ancho de banda h es muy critica en la estimación de la función de densidad mediante Parzen (Duin, 1976). En efecto, valores pequeños de h producen estimas con picos significativos centrados sobre las muestras, correspondiéndose con la posición de los núcleos sobre dichas muestras. Por el contrario, cuando h es muy grande cada patrón realiza la misma contribución en cada punto \mathbf{x} ; el resultado, en este caso, es una sobre-estima de $p_p(\mathbf{x} / w_j)$. Como se indica en Duin (1976) el ancho de la ventana debe elegirse de acuerdo con el número de muestras procesadas. Una discusión más en profundidad sobre la elección del ancho de la ventana se puede encontrar en los trabajos de Jain y Ramaswami (1988) o Parzen (1962). En efecto, en el primero aparece un detallado estudio para el caso bi-dimensional con un núcleo Gaussiano para valores de n_j y h . En dicho estudio se encuentran como valores aceptables los obtenidos con $n_j = 256$ y $h = 1$, si bien los mejores resultados se obtienen para $n_j \rightarrow +\infty$ sin importar el valor de h . A partir de aquí y con el fin de determinar el mejor valor de h , en este trabajo se ha diseñado una estrategia, que se concreta en los siguientes pasos,

1. Fijar n_j a 256 y h igual a 1.
2. Variar r desde 0.0 hasta 0.5 con incrementos de 0.05.
3. Obtener los valores de h_j en función de los valores de r según la ecuación (3.24)

4. Como durante la fase de entrenamiento el número de muestras disponibles resulta conocido, siendo exactamente n_j , se vuelve a variar r como antes desde 0.0 hasta 0.5 con incrementos de 0.05; de este modo, por medio de la ecuación (3.24) se obtiene una secuencia de valores de $h = h_j n_j^{r/q}$ donde para cada r , h_j es el obtenido en el paso tres.
5. La única cuestión pendiente consiste en determinar el mejor valor de r y por tanto de h , ya que n_j viene fijado por el número de muestras de la clase y h_j se calcula mediante los valores obtenidos en el paso tres; el mejor valor de r buscado se determina a partir de los datos disponibles mediante validación cruzada (Duda y col., 2001), tal y como se explica en la sección 5.3.1.2.

En la figura 3.2 se representa la variación de h_j frente a r par los valores de $h = 1$ y $n_j = 256$.

De acuerdo con la ecuación (3.23), este clasificador estima la función de densidad de probabilidad teniendo en cuenta las muestras de entrenamiento asociadas a cada clase, necesitando que las muestras estén ya distribuidas en ellas, o lo que es equivalente, la partición debe estar disponible; además utiliza las matrices de covarianza asociadas a cada una de las clases. La partición completa y las matrices de covarianza son los parámetros que este clasificador almacena en la *BC*, figura 3.1. En realidad, las matrices de covarianza son las mismas que las obtenidas por el clasificador de máxima verosimilitud, no siendo necesario su cálculo nuevamente.

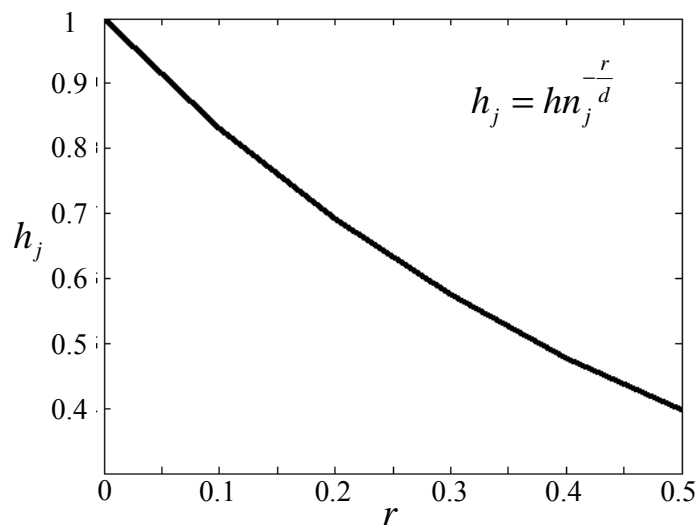


Figura 3.2 Variación de h_j frente a r para valores dados de h y n_j

3.3.3.2. Fase de decisión

Durante la fase de decisión el clasificador de Parzen extrae de la BC tanto las matrices de covarianza C_j como las muestras de entrenamiento disponibles y distribuidas en sus correspondientes clases. Con ellas se genera la función de densidad de probabilidad dada en (3.23). De este modo, ante la llegada de una nueva muestra \mathbf{x}_s se obtienen las probabilidades condicionales según dicha ecuación, $p_p(\mathbf{x}_s | w_j)$. La probabilidad de que la muestra \mathbf{x}_s pertenezca a la clase w_j puede obtenerse aplicando de nuevo la regla de Bayes,

$$P_p(w_j | \mathbf{x}_s) = \frac{p_p(\mathbf{x}_s | w_j) P(w_j)}{\sum_{j=1}^c p_p(\mathbf{x}_s | w_j)} \quad (3.25)$$

Al igual que en el caso de la estimación por máxima verosimilitud, es necesario conocer la probabilidad a priori. No obstante, en el supuesto de que no se conozca se asume que toma el mismo valor para todas las clases, por ejemplo 0.5; esta es la suposición hecha en este trabajo de investigación, al igual que en el caso del estimador de máxima verosimilitud, como ya se indicó previamente.

La clasificación de \mathbf{x}_s se determina según la siguiente regla de decisión (Duda y col., 2001),

$$\mathbf{x}_s \in w_j \quad \text{si} \quad P(w_j | \mathbf{x}_s) > P(w_k | \mathbf{x}_s) \quad \forall k \neq j \quad (3.26)$$

Debido a que el denominador en la ecuación (3.25) es el mismo para todas las clases, éste no interviene en la decisión, por lo que la regla (3.26) puede reescribirse como sigue,

$$\mathbf{x}_s \in w_j \quad \text{si} \quad p(\mathbf{x}_s | w_j) P(w_j) > p(\mathbf{x}_s | w_k) P(w_k) \quad \forall k \neq j \quad (3.27)$$

Como en el estimador de máxima verosimilitud, cuando las probabilidades a priori no se conocen de antemano, siendo este nuestro caso, la regla de decisión resulta ser finalmente,

$$\mathbf{x}_s \in w_j \quad \text{si} \quad p_p(\mathbf{x}_s | w_j) > p_p(\mathbf{x}_s | w_k) \quad \forall k \neq j \quad (3.28)$$

Al igual que en el caso del estimador de máxima verosimilitud y debido a la necesidad de restringir los valores de probabilidad al rango $[0, +1]$ cuando el clasificador actúa combinado con otros clasificadores, la regla de decisión utilizada es la definida en (3.26).

Las salidas del clasificador son las probabilidades dadas por la ecuación (3.25), esto es $P_p(w_j/x_s)$, donde el subíndice p las distingue de las probabilidades de salida proporcionadas por el clasificador basado en máxima verosimilitud, que se distinguen con el subíndice b , como se ha mencionado previamente.

3.3.4. Mapas Auto-Organizativos

3.3.4.1. Fase de entrenamiento

Es uno de los métodos más populares en redes neuronales utilizado para la reducción de la dimensionalidad de los datos por su capacidad para representar éstos por vectores significativos. Este método se conoce en la terminología inglesa como Self-Organizing Maps (SOM). En este trabajo utilizaremos estas siglas por su amplia difusión en la comunidad científica. Como otros muchos modelos en el campo de las redes neuronales fue propuesto originalmente como una aplicación inspirada en un fenómeno biológico por Kohonen (1989, 1990, 1997). Su descripción detallada se puede encontrar en Pajares y Cruz (2007) o Pandya y Macy (1996).

El modelo de Kohonen posee una arquitectura que consta de dos capas, una capa de entrada y una capa de Kohonen, también denominada mapa o capa de salida. Dichas capas se encuentran totalmente inter-conectadas. Cada neurona en la capa de entrada tiene una conexión hacia cada neurona en la capa de salida. En la figura 3.3 se ilustra la arquitectura de este tipo de redes en el caso bi-dimensional.

Cada una de las conexiones entre una neurona i de la capa de entrada y una neurona j de la capa de salida tiene asociado un peso, denominado peso de conexión c_{ij} , de forma que la componente x_i del vector de entrada $\mathbf{x} \in \Re^q$ es inyectada a través de i , que está conectada a j mediante la conexión y su peso. Esto significa que cada neurona j de la capa de salida tiene asociado un vector de pesos $\mathbf{c}_j \equiv \{c_{1j}, \dots, c_{ij}, \dots, c_{qj}\}$ cuyas componentes son los pesos de conexión con las neuronas de la capa de entrada. Estos pesos son los que se actualizan durante el proceso de entrenamiento, siendo el objetivo del aprendizaje.

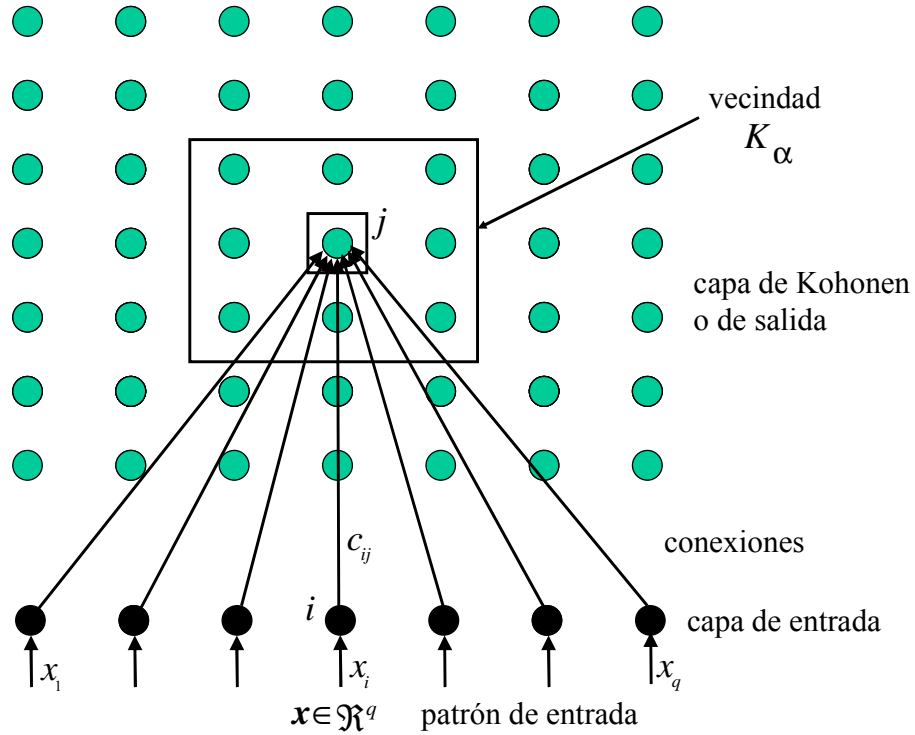


Figura 3.3 Modelo de arquitectura de Kohonen en el caso bi-dimensional

Dada una muestra de entrada \mathbf{x} , el método determina la neurona vencedora de la capa de salida. El concepto de neurona vencedora se asocia, en este caso, con el vector de pesos \mathbf{c}_j que mejor se aproxime a la muestra de entrada.

Desde el punto de vista de las redes competitivas, se diría que es la neurona vencedora la que actualiza los pesos de conexión. No obstante, el procedimiento introducido en este algoritmo no considera en exclusividad neuronas vencedoras sino que dependiendo del grado de proximidad del patrón de entrada a las distintas neuronas, así será el grado de actualización de los correspondientes pesos de conexión.

El concepto de proximidad se establece mediante la *función de vecindad* dada por la siguiente expresión,

$$K_{\alpha(t)}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\alpha^2(t)}\right) \quad \text{con} \quad \alpha(t) = \alpha_{\text{inicial}} \left(\frac{\alpha_{\text{final}}}{\alpha_{\text{inicial}}}\right)^{t/t_{\text{max}}} \quad (3.29)$$

donde t es el número de iteración y t_{max} es el número máximo de iteraciones, que se debe especificar de antemano. El ancho inicial de la vecindad α_{inicial} se elige de modo que la propia vecindad cubra todas las neuronas y el ancho final de la vecindad α_{final} controla

el grado de variación de la función de vecindad entre iteraciones consecutivas. Estos parámetros permanecen invariables una vez fijados. Se dice que \mathbf{x} está en la vecindad de \mathbf{x}' según el valor que tome dicha función, de suerte que fijado un determinado umbral, los vectores con valores por debajo del mismo se consideran que no pertenecen a la región de vecindad y viceversa. No obstante, en este trabajo no se aplica este criterio; por tanto todas las muestras mantienen un cierto grado de pertenencia a una vecindad dada, por pequeño que éste sea y por consiguiente no es necesario fijar dicho umbral.

El algoritmo de auto-organización se sintetiza de acuerdo con los pasos siguientes (Pandya y Macy, 1996),

1. *Inicialización*: inicializar el vector de pesos \mathbf{c}_j ($t = 0$), que bien pudiera ser a valores aleatorios o conocidos, siempre y cuando en este último caso, se hayan obtenido mediante algún otro procedimiento. Inicializar la razón de aprendizaje $\gamma(t = 0)$. Definir los valores involucrados en $\alpha(t)$, a saber: $\alpha_{inicial}$, α_{final} y t_{max} .
2. Para cada muestra representada por su vector \mathbf{x} realizar los pasos 3 y 4.
3. *Medida de similitud*: seleccionar la neurona j cuyo vector de pesos esté más próximo al patrón de entrada según el criterio definido en (3.30), donde $\|\cdot\|$ define una norma, elegida como la euclídea en este trabajo,

$$j \text{ vence si } \|\mathbf{x} - \mathbf{c}_j\| < \|\mathbf{x} - \mathbf{c}_k\| \quad \forall k \neq j \quad (3.30)$$

4. *Entrenamiento*: actualizar los pesos de las neuronas dentro de la región de vecindad de acuerdo con la siguiente ley,

$$\mathbf{c}_j(t+1) = \begin{cases} \mathbf{c}_j(t) + \gamma(t)[\mathbf{x} - \mathbf{c}_j(t)] & j \in K_{\alpha(t)} \\ \mathbf{c}_j(t) & \text{de cualquier otro modo} \end{cases} \quad (3.31)$$

5. Actualizar la razón de aprendizaje, según la ley definida con tal propósito, como se verá posteriormente.
6. Disminuir el valor de la función de vecindad mediante la variación de $\alpha(t)$.
7. Verificar la condición de convergencia a partir del criterio definido en (3.32), según el cual cuando la variación de todos los pesos entre dos iteraciones

consecutivas está por debajo de un determinado umbral ε , según la norma euclídea, el proceso se detiene,

$$\|c_j(t+1) - c_j(t)\| < \varepsilon \quad \forall j \quad (3.32)$$

La ley de actualización de pesos tanto para las neuronas vencedoras como para las que no han vencido se expresa como sigue,

$$c_j(t+1) = c_j(t) + \gamma(t) K_{\alpha(t)}(x, c_j(t)) [x - c_j(t)] \quad (3.33)$$

Obsérvese cómo la función de vecindad y la razón de aprendizaje determinan conjuntamente el valor de la fuerza ejercida por x sobre c_j para la actualización de éste.

Los valores de los parámetros involucrados en este algoritmo se determinan, como en el caso de los anteriores, mediante validación cruzada como se verá en el capítulo cinco, sección 5.3.1.3. Finalmente, la razón de aprendizaje se ha establecido como $\gamma(t) = 1/(10+t)$, que determina un decremento de su valor a medida que el número de iteraciones aumenta; lo cual está en concordancia con el hecho de que a mayor experiencia en la vida menor es el nivel de aprendizaje obtenido.

La aplicación del algoritmo SOM al caso de la clasificación de las texturas de las imágenes naturales requiere la definición de los patrones de entrada así como de los pesos de conexión y el número de neuronas en la capa de Kohonen. De este modo, los vectores de entrada son los vectores $x \in \Re^3$ que caracterizan las componentes espectrales de los píxeles según el modelo de color RGB adoptado en nuestro caso. Por tanto, el número de neuronas en la capa de entrada es tres, que se corresponden con cada una de las tres componentes R, G y B utilizadas.

Por otra parte, el número de neuronas en la capa de salida viene determinado por el número de clases existentes, donde el vector de pesos de conexión de cada neurona representa el centro de cada clase, es decir c_j para la clase w_j . Estos vectores de pesos o centros son los parámetros de aprendizaje que se almacenan en la BC tras el proceso de entrenamiento, según el esquema de la figura 3.1.

3.3.4.2. Fase de decisión

Como siempre, el objetivo de la fase de decisión consiste en clasificar una nueva muestra \mathbf{x}_s en alguna de las clases w_j existentes. Con tal propósito, se recuperan de la BC los centros (pesos) \mathbf{c}_j almacenados durante la fase de entrenamiento, determinándose la proximidad de la muestra a todos los centros de las clases. La proximidad se establece en función de un criterio de mínima distancia. Esta distancia debe ser la misma que la utilizada en la fase de entrenamiento, por tanto la euclídea. La clasificación de la muestra se lleva a cabo según la regla de decisión siguiente,

$$\mathbf{x}_s \in w_j \quad \text{si} \quad \|\mathbf{x}_s - \mathbf{c}_j\| < \|\mathbf{x}_s - \mathbf{c}_k\| \quad \forall k \neq j \quad (3.34)$$

Las salidas proporcionadas por este clasificador son las distancias euclídeas entre \mathbf{x}_s y cada uno de los centros \mathbf{c}_j , es decir $d(\mathbf{x}_s, \mathbf{c}_j) = \|\mathbf{x}_s - \mathbf{c}_j\|$.

3.3.5. Cuantización vectorial

Bajo esta categoría se describe el método de cuantización vectorial, conocido en la literatura inglesa como *Learning Vector Quantization* (LVQ). Por su amplia difusión en la comunidad científica, nos referimos a él con estas siglas. En este trabajo durante la fase de entrenamiento se consideran dos versiones del mismo, a saber el original LVQ1 (Kohonen, 1989) y una de sus variantes conocida como LVQ2 (Kohonen, 1990). En cualquiera de los dos casos se trata de sendos métodos de aprendizaje supervisado, que suponen una extensión de los métodos basados en la arquitectura de Kohonen, introducida previamente. También, en ambos casos es necesario especificar el número de clases en las que se van a clasificar las muestras, así como la distribución de éstas en sus respectivas clases, es decir, se debe especificar la partición.

3.3.5.1. Fase de entrenamiento

El proceso de entrenamiento para los dos métodos LVQ1 y LVQ2 sigue los mismos pasos que el esquema del SOM (Kohonen 1989, 1997, Kosko 1992), con la única excepción del paso cuatro, ya que en este caso, los pesos se modifican de forma distinta según la ecuación (3.35).

$$\mathbf{c}_j(t+1) = \begin{cases} \mathbf{c}_j(t) + \gamma(t)[\mathbf{x} - \mathbf{c}_j(t)] & \text{si } \mathbf{c}_j \text{ vence correctamente con } \mathbf{x} \\ \mathbf{c}_j(t) - \gamma(t)[\mathbf{x} - \mathbf{c}_j(t)] & \text{si } \mathbf{c}_j \text{ vence incorrectamente con } \mathbf{x} \\ \mathbf{c}_j(t) & \text{de otro modo} \end{cases} \quad (3.35)$$

Obsérvese el concepto de aprendizaje supervisado inmerso en las condiciones de victoria correcta o incorrecta. La ecuación (3.35) expresa, en la primera línea, un movimiento de aproximación del centro \mathbf{c}_j hacia \mathbf{x} , debido al correcto vencimiento de \mathbf{c}_j y en la segunda línea un alejamiento del centro \mathbf{c}_j respecto de \mathbf{x} por su incorrecto vencimiento. La tercera línea expresa que el resto de centros \mathbf{c}_i permanecen inalterados. Por vencer correctamente se entiende que si \mathbf{x} pertenece a la clase w_j entonces se cumple que el centro más próximo es \mathbf{c}_j , es decir el centro de su clase. Por el contrario, vencer incorrectamente con \mathbf{x} se refiere a que el centro más próximo a \mathbf{x} es \mathbf{c}_j , si bien \mathbf{x} no pertenece a la clase w_j .

La variante LVQ2 es una extensión de la LVQ1 de modo que en la primera se actualizan tanto el campeón como el subcampeón bajo ciertas condiciones. La idea principal es que cuando la neurona (centro) vencedora no represente la categoría correcta, y la subcampeona sí, se actualicen ambas. En este caso, la vencedora se aleja de la muestra \mathbf{x} y la subcampeona se aproxima, o dicho de otra manera, el centro vencedor es castigado y el subcampeón recompensado. El método impone una restricción adicional en el sentido de que la distancia d_c entre \mathbf{x} y el centro vencedor no sea muy diferente de la distancia d_r entre \mathbf{x} y el centro subcampeón. Cuando no se cumpla esta restricción, el entrenamiento se realiza exactamente igual que en el caso del LVQ1.

La condición de que d_r y d_c no sean muy diferentes se expresa en términos de un ancho de ventana dentro de la cual, la condición se considera cumplida. La ventana se define según la ecuación (3.36).

$$\frac{d_c}{d_r} = 1 - \sigma \quad \text{y} \quad \frac{d_r}{d_c} = 1 + \sigma \quad (3.36)$$

Cuando las dos condiciones dadas en (3.36) se cumplen, el entrenamiento se realiza de acuerdo con las siguientes expresiones según corresponda al centro subcampeón o campeón respectivamente,

$$\text{Subcampeón: } \mathbf{c}_j(t+1) = \mathbf{c}_j(t) + \gamma(t) [\mathbf{x} - \mathbf{c}_j(t)] \quad (3.37)$$

$$\text{Campeón: } \mathbf{c}_j(t+1) = \mathbf{c}_j(t) - \gamma(t) [\mathbf{x} - \mathbf{c}_j(t)]$$

Desde el punto de vista notacional, conviene señalar que aunque tanto en LVQ1 como en LVQ2 y en SOM los centros se expresan como \mathbf{c}_j se refieren a centros diferentes, debido a las distintas leyes de actualización de los mismos, distinguiéndolos implícitamente por su asociación a cada uno de los métodos.

El criterio de convergencia en LVQ1 y LVQ2 es el mismo que el utilizado en el algoritmo SOM, de suerte que cuando se consigue la convergencia o se llega a un número máximo de iteraciones t_{max} , la actualización se detiene, en cuyo caso, los últimos centros actualizados se almacenan en la BC según la figura 3.1. La estimación de los parámetros involucrados en este proceso se lleva a cabo mediante correlación cruzada, tal y como se describe en el capítulo cinco, sección 5.3.1.4, siguiendo las mismas pautas que para el resto de métodos.

3.3.5.2. Fase de decisión

La fase de decisión es idéntica a la del algoritmo SOM, con la salvedad de que los centros utilizados son los obtenidos durante el entrenamiento de los respectivos métodos LVQ1 y LVQ2.

Como en el caso del SOM, las salidas proporcionadas por ambos clasificadores son también las distancias euclídeas entre \mathbf{x}_s y cada uno de los centros \mathbf{c}_j , es decir, $d(\mathbf{x}_s, \mathbf{c}_j) = \|\mathbf{x}_s - \mathbf{c}_j\|$.

3.3.6. Resumen de clasificadores individuales

La tabla 3.1 muestra una síntesis de los seis clasificadores individuales descritos previamente desde el punto de vista de ambas fases de entrenamiento y decisión. Respecto de la fase de entrenamiento se indica su naturaleza, los parámetros que aprende o estima durante dicha fase y los parámetros que son necesarios fijar para llevar a cabo el proceso. Respecto de la fase de decisión se indica tanto la regla de decisión cuando el clasificador actúa de forma independiente como las salidas que proporciona para su posterior combinación con otros clasificadores.

Tabla 3.1 Características generales de cada clasificador en las fases de aprendizaje y decisión.

Clasificador	Fase de Aprendizaje			Fase de Decisión	
	Natura- leza	Parámetros aprendidos	Parámetros requeridos (3)	Regla de decisión	Salidas fase de decisión
Agrupamiento borroso	Supervi- sado	centros de las clases \mathbf{v}_j	<ul style="list-style-type: none"> • m (peso exponencial) • t_{max} (máx. número de iteraciones) • \mathcal{E} (criterio de convergencia) 	máximo grado de pertenencia	Grados de pertenencia de \mathbf{x}_s a las clases w_j : μ_{sj}
Estima de máxima verosimilitud	No supervi- sado (1)	centros de las clases y matrices de covarianza \mathbf{m}_j, C_j	no aplicable	máxima probabili- dad	Probabilidades de pertenencia de \mathbf{x}_s a las clases w_j : $P_b(w_j/\mathbf{x}_s)$
Ventana de Parzen	Supervi- sado	Función de densidad de probabilidad: matices de covarianza C_j y muestras de entrenamiento	<ul style="list-style-type: none"> • r (coeficiente del ancho de banda) • h (parámetro de suavizado) 	máxima probabili- dad	Probabilidades de pertenencia de \mathbf{x}_s a las clases w_j : $P_p(w_j/\mathbf{x}_s)$
Mapas auto- organizativos	No Supervi- sado (2)	Centros de las clases \mathbf{c}_j	<ul style="list-style-type: none"> • $\gamma(t=0)$ (razón de aprendizaje), • $\alpha_{inicial}, \alpha_{final}$ • t_{max} (máx. número de iteraciones) • \mathcal{E} (criterio de convergencia) 	mínima distancia	Distancias entre \mathbf{x}_s y los centros de las clases \mathbf{c}_j : $d(\mathbf{x}_s, \mathbf{c}_j)$
Cuantización vectorial LVQ1	Supervi- sado	Centros de las clases \mathbf{c}_j	<ul style="list-style-type: none"> • $\gamma(t=0)$ (razón de aprendizaje) • t_{max} (máximo número de iteraciones) • \mathcal{E} (criterio de convergencia) 	mínima distancia	Distancias entre \mathbf{x}_s y los centros de las clases \mathbf{c}_j : $d(\mathbf{x}_s, \mathbf{c}_j)$
Cuantización vectorial LVQ2	Supervi- sado	Centros de las clases \mathbf{c}_j	<ul style="list-style-type: none"> • $\gamma(t=0)$ (razón de aprendizaje) • t_{max} (máx. número de iteraciones) • \mathcal{E} (criterio de convergencia) • σ (ancho ventana) 	mínima distancia	Distancias entre \mathbf{x}_s y los centros de las clases \mathbf{c}_j : $d(\mathbf{x}_s, \mathbf{c}_j)$

(1) La estimación mediante máxima verosimilitud es No supervisada en sí misma (Duda y col., 2001), si bien desde el momento en el que se utilizan las muestras distribuidas en clases, se pierde esta categoría.

(2) Tradicionalmente, el SOM se ha considerado no supervisado, si bien por el mismo razonamiento anterior, al tomar la partición como referencia pierde esta categoría.

(3) El procedimiento concreto para estimar dichos parámetros se describe en el capítulo cinco, sección 5.3.

3.4. Funciones criterio para la partición

Para estimar los parámetros derivados del proceso de aprendizaje, los seis métodos expuestos en la sección 3.3 necesitan la distribución en c clases, w_1, w_2, \dots, w_c de las n muestras contenidas en $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \in \mathcal{R}^q$; esto es, la partición de dichas muestras. Dicha distribución puede realizarse de dos formas, una es manualmente bajo la supervisión de un experto; la otra es de forma automática sin necesidad de supervisión. Esta última es la adoptada en este trabajo de investigación con el fin último de conseguir la automatización del proceso de aprendizaje, que constituye uno de los objetivos de la investigación.

Se plantea ahora el problema de hasta qué punto la partición de las muestras sin supervisión puede considerarse como válida. En esta sección se aborda precisamente esta problemática. Con tal propósito, en primer lugar se establecen las premisas previas, que determinan cuándo una partición se considera válida y en segundo lugar, cómo se mide el grado de validez de dicha partición.

Resulta comúnmente aceptado que cuanto más similares son entre sí las muestras de una misma clase y a la vez más diferencias existen entre las muestras de clases distintas, tanto mejor será la partición. Por tanto, el objetivo consiste en obtener una partición que cumpla lo mejor posible las dos premisas anteriores.

La validación de las particiones ha sido un tema ampliamente tratado en la literatura, como puede deducirse de los estudios realizados por Volkovich y col. (2008) y referencias asociadas. A continuación abordamos esta cuestión, teniendo en cuenta la vinculación de los métodos de validación con los clasificadores utilizados.

Uno de los clasificadores es el de *agrupamiento borroso*, que estima los grados de pertenencia de las muestras a las clases, ecuación (3.3). Pues bien, dado que la función criterio conocida como *coeficiente de partición* permite obtener un coeficiente de validez según dichos grados de pertenencia (Zimmermann, 1991), parece lógico considerar tal función para la validación de la partición. Por otro lado, el estimador paramétrico basado en *máxima verosimilitud* estima la media de los datos en cada clase, además los *mapas auto-organizativos* y los dos métodos de *cuantización vectorial* estiman centros de las clases según sus correspondientes procedimientos. Dichos centros, al igual que las

medias son representantes de las clases. Por esta razón, una función criterio para validar la partición desde esta perspectiva puede ser la conocida como *suma del error al cuadrado*, que utiliza representantes de las clases como referencia (Duda y col., 2001). Finalmente, el estimador no paramétrico de la *ventana de Parzen* establece una función de densidad para cada clase en función de las muestras de la clase. Por este motivo, la función criterio conocida como *varianza mínima relacionada* (Duda y col., 2001), puede ser apropiada en este caso.

Por otro lado, existen aún dos cuestiones útiles relacionadas con las funciones criterio que pueden ser aprovechadas convenientemente. En primer lugar, como se verá posteriormente durante el estudio de los clasificadores combinados, uno de los problemas que surgen consiste en determinar cuál es la importancia relativa de cada clasificador individual dentro de la combinación, es decir cuál de ellos presenta un mejor o peor comportamiento. En este trabajo se ha diseñado un procedimiento para deducir el comportamiento de los diferentes clasificadores individuales en función de los valores proporcionados por las funciones criterio. En segundo lugar, los clasificadores individuales estudiados hasta ahora son de naturaleza supervisada, bien por su propia definición o por aproximación, ver tabla 3.1. Pues bien, gracias al comportamiento de las funciones criterio, es posible automatizar el proceso y conseguir un diseño no supervisado.

En las tres sub-secciones siguientes se definen cada uno de las tres funciones criterio, mencionadas previamente, para la validación. En la cuarta sub-sección se analiza el comportamiento de estas funciones según el número de clases. A partir de este comportamiento se obtiene finalmente, en la sub-sección quinta, la función de validación de la partición, que permitirá determinar el número de clases a considerar como óptimo.

3.4.1. Coeficiente de Partición

El *coeficiente de partición* (PC) es una función criterio que se define como sigue (Zimmermann, 1991; Kim y col., 2003):

$$PC(U; c) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^2 \quad \text{con} \quad \frac{1}{c} \leq PC(U; c) \leq 1 \quad (3.38)$$

donde U es la matriz de grados de pertenencia μ_{ij} a las clases w_j , definida en la sección 3.3.1.1; n es el número de muestras y c el número de clases. Obsérvese que el dominio de PC se reduce al intervalo $[1/c, 1]$. Según este coeficiente, las mejores particiones se obtienen para valores máximos de PC , como se analizará posteriormente.

Existen otras funciones basadas en los grados de pertenencia tales como *entropía de la partición* o *proporción del exponente*, entre otras, pudiéndose encontrar en Kim y col., (2003); si bien, su comportamiento no introduce ninguna aportación relevante con respecto al coeficiente de partición, motivo por el cual no se han considerado.

3.4.2. Suma del error al cuadrado

La *suma del error al cuadrado* (SE) es otra de las funciones criterio ampliamente utilizadas para la validación de particiones de datos (Duda y col. 2001). Dado el número de muestras n_j pertenecientes a la clase w_j y la media de dichas muestras \mathbf{m}_j en esa clase, ecuación (3.39), la función criterio SE queda definida según la ecuación (3.40),

$$\mathbf{m}_j = \frac{1}{n_j} \sum_{\mathbf{x} \in w_j} \mathbf{x} \quad (3.39)$$

$$SE(\mathbf{m}_j, \mathbf{x}, c) = \sum_{j=1}^c \sum_{\mathbf{x} \in w_j} \|\mathbf{x} - \mathbf{m}_j\|^2 \quad (3.40)$$

En la función definida por la ecuación (3.40) el vector media \mathbf{m}_j puede sustituirse por los correspondientes centros \mathbf{c}_j obtenidos mediante los mapas auto-organizativos y los dos cuantizadores vectoriales. Esto es lo que se hace realmente cuando se utiliza la función criterio SE en relación con estos tres métodos de clasificación.

La función SE tiene una interpretación muy simple. En efecto, dada una clase w_j el vector media es el mejor representante de la clase para las muestras en w_j . A partir de la ecuación (3.40) se deduce que si las muestras en las clases se encuentran muy agrupadas alrededor de sus respectivas medias, SE toma valores mínimos, llegando a ser cero en el hipotético caso de que todas las muestras coincidan exactamente con sus medias. Por tanto, SE mide el error al cuadrado total de la distribución de las n muestras $\mathbf{x}_1, \dots, \mathbf{x}_n$ con respecto a las c medias o centros de las clases $\mathbf{m}_1, \dots, \mathbf{m}_c$. Es evidente, que el valor de SE

depende del agrupamiento de las muestras en las clases y del número de clases; la partición óptima se define como la que minimiza el coeficiente SE . Las agrupaciones de este tipo, a menudo, se denominan particiones de *mínima varianza*, debido a que, de alguna manera se miden dispersiones de las muestras con respecto a las medias o centros.

Ahora se plantea el problema de qué tipo de particiones son deseables bajo la perspectiva de esta función. Básicamente, la función SE proporciona un criterio adecuado cuando las clases forman agrupaciones compactas, estando además bien separadas unas de otras, es decir, sin solapamiento entre ellas.

Un problema que plantea esta función surge porque determinadas agrupaciones naturales se ven perjudicadas, mientras otras agrupaciones antinaturales se ven claramente favorecidas. En la figura 3.4 se muestra un sencillo ejemplo poniendo de manifiesto lo anterior. En la parte superior aparecen dos clases bien separadas, donde las muestras están agrupadas en dos clases de forma natural, sin embargo el valor de SE es mayor que el obtenido con la agrupación que aparece en la parte inferior, donde la agrupación resulta ser más bien antinatural y además con un alto grado de solapamiento entre ambas clases.

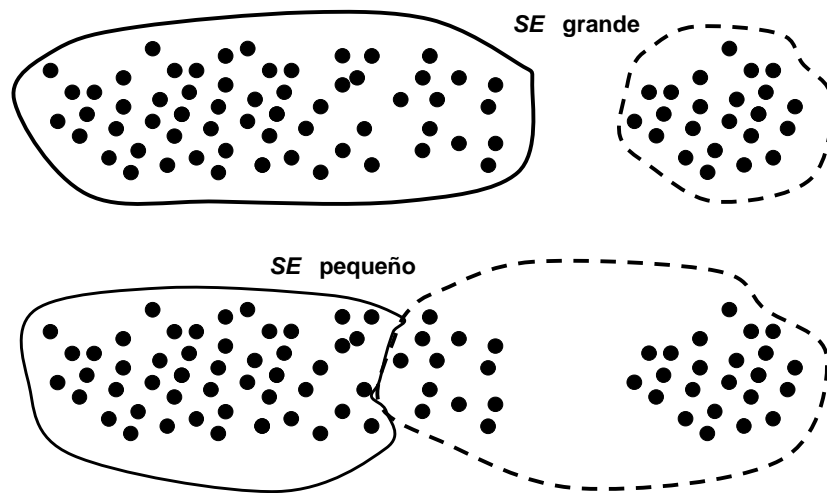


Figura 3.4 Dos agrupamientos con diferentes valores del coeficiente SE

En la figura 3.5 se muestra un ejemplo de agrupamiento para una de las imágenes utilizadas en los experimentos llevados a cabo en el presente trabajo, mostrándose la distribución de las mismas en cuatro clases, así como la posición de las medias obtenidas mediante la ecuación (3.39). La proyección se ha realizado en el espacio bidimensional a través de la técnica basada en la transformada de *Componentes Principales* (Pajares y

Cruz, 2007a). En este ejemplo, las clases se diferencian perfectamente y las muestras de cada clase aparecen agrupadas alrededor de sus medias, aunque bien es cierto que con un cierto grado de dispersión. Este es el comportamiento general de las muestras procedentes de las imágenes utilizadas en los distintos experimentos realizados bajo la presente investigación. Por tanto, no se da la situación problemática expuesta previamente respecto del comportamiento de la función SE , lo que permite elegir esta función SE para la validación de las particiones.

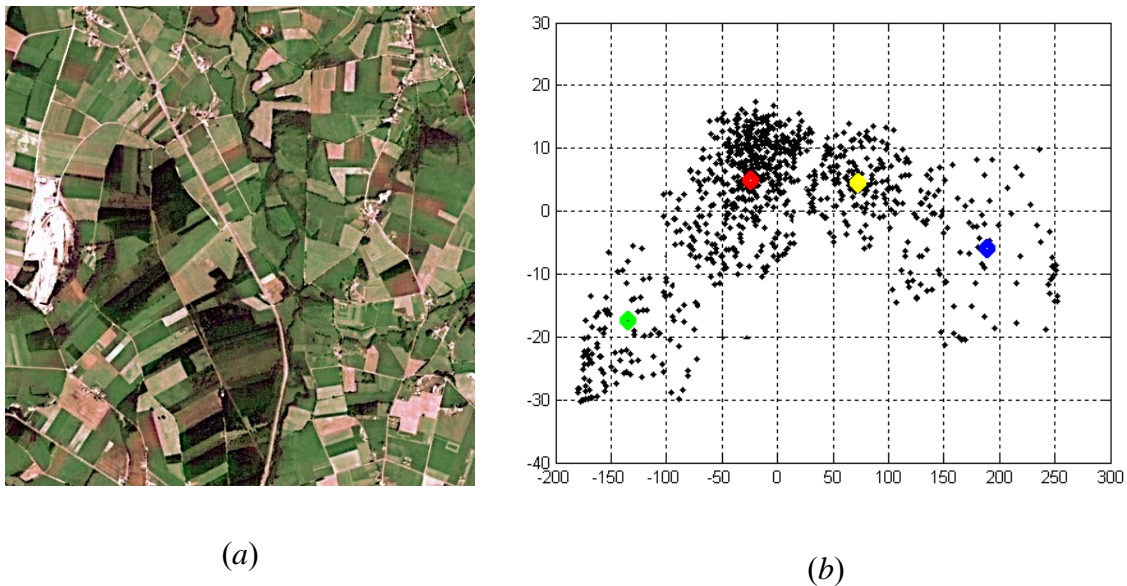


Figura 3.5 (a) Imagen original, (b) distribución de las muestras en cuatro clases.

En Duda y col. (2001) se describen otras funciones criterio basadas en las medias de las clases para la validación de particiones, entre las que destacan las derivadas de las matrices de dispersión (“scatter matrices”), que no se contemplan en este trabajo de investigación por no realizar ninguna aportación de relevancia con respecto a la función SE . Lo mismo puede afirmarse respecto de otras funciones tales como la distancia de Jeffries-Matusita (Pajares y Cruz, 2007a).

3.4.3. Criterio de varianza mínima relacionada

El *criterio de varianza mínima relacionada* (VC) (Duda y col. 2001), en lugar de utilizar como referencia las medias de las clases, emplea todas las muestras. Es decir, considera la dispersión mutua de las muestras en las clases según la ecuación (3.41).

$$VC(\mathbf{x}, c) = \frac{1}{2} \sum_{j=1}^c n_j \bar{s}_j \quad \text{con} \quad \bar{s}_j = \frac{1}{n_j} \sum_{\mathbf{x} \in w_j} \sum_{\mathbf{x}' \in w_j} \|\mathbf{x} - \mathbf{x}'\|^2 \quad (3.41)$$

En la ecuación anterior, \bar{s}_j se interpreta como la media de las distancias al cuadrado entre las muestras en la clase w_j y pone el énfasis en el hecho de que el criterio de la suma de los errores al cuadrado utiliza la distancia euclídea como medida de similitud. En lugar de la distancia Euclídea se podría utilizar alguna otra medida de similitud, tal como por ejemplo la media, mediana, o simplemente la máxima distancia de las muestras en la clase. No obstante, en este trabajo se ha adoptado la función criterio VC en su versión original, tal y como se define en la ecuación (3.41). Como en el caso del SE , la partición óptima se define como la que minimiza VC . Lo dicho anteriormente para SE , en relación a su comportamiento respecto de particiones como las mostradas en las figuras 3.4 y 3.5, es también aplicable a VC .

3.4.4. Comportamiento de las funciones criterio

La función criterio PC es monótona creciente, mientras que SE y VC son monótonas decrecientes, en los tres casos en relación al número de clases; llegando a alcanzar sus valores extremos exactamente cuando el número de muestras es igual al número de clases ($c = n$). En el caso de PC dicho valor extremo es la unidad, que coincide con el máximo; mientras que para SE y VC los extremos son los valores mínimos, que se corresponden con el cero. Obviamente la obtención de una partición con un número de clases igual al número de muestras, donde cada clase posee una única muestra, no tiene ningún sentido a efectos prácticos. Por este motivo, el objetivo se centra ahora en obtener un criterio para determinar bajo qué condiciones de máximo o mínimo se puede considerar válida la partición.

Del estudio teórico sobre el comportamiento de las funciones anteriores realizado en Zimmermann (1991), Duda y col. (2001) o Kim y col., (2003), así como de su aplicación a diferentes tipos de datos como los descritos en la sección 5.2, se deduce que tanto SE como VC disminuyen rápidamente hasta alcanzar el número de clases correcto ($c = \hat{c}$), según la información proporcionada por las fuentes que suministran los datos, disminuyendo mucho más lentamente a partir de aquí hasta alcanzar el mínimo absoluto.

PC crece según el número de clases, si bien lo hace mucho más suavemente y sin cambios bruscos, que el decrecimiento de SE y VC .

Según lo anterior, uno de los procedimientos más habituales para estimar el mejor número de clases, esto es la mejor partición, consiste en determinar para qué número de clases se produce ese cambio de tendencia en el descenso de las funciones criterio SE y VC y en el ascenso de PC , teniendo en cuenta sus valores para el número de clases $c + 1$ y el valor para c , esto es lo que refleja la ecuación (3.42). No obstante, debido a que el rango de variación de los valores de las funciones criterio es muy diferente entre ellas, para analizar su comportamiento relativo, lo más apropiado consiste en normalizar dichas variaciones en el rango $[0,1]$, esto es lo que se realiza en la ecuación (3.43).

$$\Delta^h(c) = h(c) - h(c+1) \text{ y } \Delta^{PC}(c) = f(c+1) - f(c); \quad c = 1, \dots, G-1 \quad (3.42)$$

$$\hat{\Delta}^f(c) = \frac{\Delta^f(c)}{\sum_c \Delta^f(c)} \quad \text{con } c = 1, \dots, G-1 \quad (3.43)$$

donde h en la ecuación (3.42) representa a SE o VC y f en la ecuación (3.43) representa a una de las tres funciones criterio definidas previamente, es decir PC , SE o VC ; G es el número máximo de clases considerado. Por esta razón, el sumatorio en el denominador varía desde 1 hasta $G - 1$. Obsérvese en la ecuación (3.42) el cómputo de las diferencias, que en el caso de SE y VC son $h(c) - h(c+1)$ por ser monótonas decrecientes y en el caso de PC es $f(c+1) - f(c)$ por su monotonía creciente.

En la figura 3.6 se representa la variación relativa de las funciones PC , SE y VC , según la ecuación (3.46), frente al número de clases para los datos representados en la figura 3.5(b). Como se ha mencionado previamente, al explicar esta última figura, se considera que el número de clases adecuado es cuatro, según el criterio del experto humano o usuario final. Conviene mencionar que para otros conjuntos de datos, como los mostrados en el capítulo cinco, tabla 5.1, que poseen distintos números de clases, el comportamiento es similar al mostrado en la figura 3.6 y por tanto, cuando se alcanza el número de clases especificado por los proveedores de los datos, la variación relativa de las funciones se suaviza considerablemente.

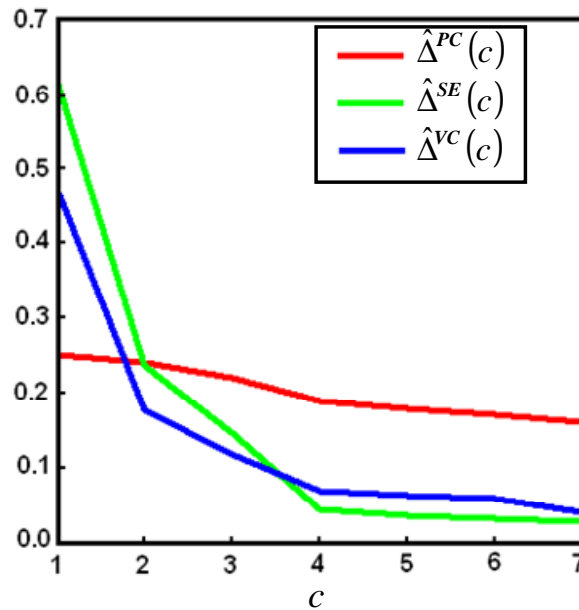


Figura 3.6 Variación relativa de los coeficientes PC , SE y VC entre números de clases consecutivos

A la vista de los resultados mostrados en la figura 3.6 se pueden realizar las siguientes observaciones de relevancia:

- La variación de PC , presenta una disminución progresiva y suave, esto es sin saltos bruscos.
- Las variaciones más pronunciadas provienen de SE y VC , siendo en general la función SE la que presenta mayores variaciones, aunque sea ligeramente, en términos absolutos.
- En general, las diferencias hasta el número cuatro, varían bruscamente; a partir de dicho número la variación se atenúa, tornándose mucho más suave.

Estas observaciones son de vital importancia, dado que serán posteriormente utilizadas como criterios, tanto para validar las particiones como para estimar las competencias relativas de los clasificadores individuales, que participan en la combinación.

El objetivo consiste en encontrar un umbral, de suerte que cuando la variación relativa se encuentre por debajo del mismo, la partición se considera válida. Los detalles concretos sobre la obtención de este umbral se posponen para la sección 5.3.1.5. El mecanismo para determinar las competencias de los clasificadores se proporciona en la sección siguiente, matizándose posteriormente en el capítulo cuatro, sección 4.2.3.

3.4.5. Obtención de las competencias de los clasificadores individuales

Una aportación importante, realizada por el trabajo de investigación que se presenta, consiste en determinar las competencias de los clasificadores individuales teniendo en cuenta el comportamiento de las funciones criterio definidas previamente. Esto permite determinar los pesos específicos con los que los clasificadores individuales intervendrán en las dos estrategias combinadas y planteadas en el capítulo cuatro.

A la vista del comportamiento observado en la figura 3.6, la mejor discriminación la proporciona el coeficiente SE , ya que cuando se llega al hipotético número óptimo de clases designado como $c = \hat{c}$ (en el caso de la figura $\hat{c} = 4$), la variación relativa de $\hat{\Delta}^{SE}(\hat{c})$ resulta ser la menor de las tres. Es decir, ante el número de clases considerado como correcto, el decrecimiento relativo de la función SE es menor que en las otras dos, lo que indica que está más cerca de su mínimo, observándose además que a partir de este momento la variación resulta ser relativamente suave.

Volviendo al principio de la sección, recordamos que PC se asocia con el clasificador borroso; SE con el estimador de máxima verosimilitud, los mapas auto-organizativos y los dos métodos de cuantización vectorial; y finalmente VC con el estimador de la ventana de Parzen. No obstante, antes de continuar, conviene aclarar que debido al peor comportamiento de PC frente a SE y VC , según se desprende del análisis de la sección previa, en este trabajo de investigación, se ha optado por prescindir de PC y asociar el clasificador borroso con SE . Esta asociación se justifica por el hecho de que este clasificador, además de los grados de pertenencia obtiene sus propios centros de clase según la ecuación (3.4), equiparándose así con los otros cuatro clasificadores que también estiman centros de las clases.

A partir de la asociación entre clasificadores y funciones criterio y teniendo en cuenta las variaciones relativas expresadas por la ecuación (3.43) para dichas funciones, se pueden determinar las *competencias o relevancias absolutas* de los clasificadores. Ambos términos se utilizarán indistintamente. Así, el clasificador más relevante será aquél con la menor variación relativa de su función criterio para el número de clases considerado correcto \hat{c} , esto es, el clasificador asociado con el menor valor de su $\hat{\Delta}^f(\hat{c})$. Puesto

que $\hat{\Delta}^f(\hat{c})$ varía en el rango $[0,1]$, la competencia absoluta para un clasificador dado se puede obtener como,

$$\alpha^f = 1 - \hat{\Delta}^f(\hat{c}) \quad (3.44)$$

donde f hace referencia, a una de las dos funciones criterio SE o VC , según se ha expuesto previamente.

Según la ecuación (3.44) el clasificador más importante, en términos absolutos, es aquél asociado con el mayor valor dado por α^f . En los capítulos cuatro, sección 4.2.3 y cinco, sección 5.5.1 se proporcionan los detalles y resultados de la aplicación de este concepto, donde a partir de las competencias absolutas se obtienen los pesos específicos asociados a cada clasificador individual y otras medidas de su importancia relativa, que serán utilizadas por las estrategias combinadas propuestas en este trabajo de investigación.

3.5. Métodos de clasificación combinados

Hasta ahora se han estudiado un conjunto de seis clasificadores clásicos catalogados como individuales, que realizan los procesos de aprendizaje y clasificación por sí mismos. En los últimos tiempos se ha abierto una línea de investigación basada en la combinación de clasificadores, son los denominados sistemas combinados o multi-clasificador, cuya finalidad estriba en mejorar los resultados de los clasificadores individuales, que participan en la combinación cuando se consideran aisladamente.

Siguiendo la filosofía expuesta en Kuncheva (2002, 2004) se plantea el dilema de si la combinación de clasificadores está o no suficientemente justificada. Ho (2002) aborda el tema de la clasificación tratando de buscar las mejores combinaciones posibles de clasificadores, aún a costa del aumento de la complejidad computacional. Volviendo a Kuncheva (2004), la combinación de clasificadores se perfila como un paso natural hacia un nivel superior de conocimiento, se parte del hecho de que previamente se dispone de una masa crítica de ese conocimiento de base suficientemente abundante. Este conocimiento proviene, a su vez, de la utilización de los clasificadores individuales en muy diversas áreas de aplicación del reconocimiento de patrones.

Dietterich (2000) sugirió tres tipos de razones por las cuales un clasificador combinado puede resultar más eficaz que un clasificador individual, a saber: *estadísticas*, *computacionales* y de *representación*.

- a) *Estadísticas*: supongamos que se dispone de un conjunto de clasificadores para llevar a cabo la tarea de clasificar un conjunto de datos X . Cada uno de ellos producirá ciertos resultados, cabe la posibilidad de que elegido uno al azar, éste no sea exactamente el óptimo para el conjunto de datos dado, con lo que la elección no habrá sido la adecuada. Por el contrario, si seleccionamos varios de ellos, la probabilidad de que la elección sea errónea disminuye, soslayando así el riesgo de elegir un clasificador no apropiado.
- b) *Computacionales*: algunos algoritmos de clasificación basados en búsquedas heurísticas pueden llegar a bloquearse por falta de solución. En los clasificadores basados en métodos de optimización se puede caer en mínimos locales difíciles de superar. Por este motivo, la combinación de diferentes clasificadores puede evitar alguna de las situaciones anteriores no deseables.
- c) *Representación*: se refiere al hecho de que determinados tipos de datos pueden adaptarse mejor a ciertos clasificadores, con el fin de proceder a su clasificación. Por ejemplo, supóngase que los datos X se clasifican muy bien con clasificadores de tipo lineal, pues bien, si utilizáramos otro tipo de clasificadores, tales como redes neuronales, que carecen de la mencionada linealidad, es posible que los resultados no fuesen los deseados. Combinando clasificadores se podrían evitar tales situaciones.

3.5.1. Terminología y taxonomías

Desde el año 2000 se viene celebrando el “International Workshop on Multiple Classifier Systems”, (MCS, 2009). También en los últimos años se han celebrado los “International Workshops on Hybrid Artificial Intelligent Systems (HAIS’05-09)”, los trabajos publicados en Guijarro y col. (2007a, 2008), fruto de la investigación desarrollada en el ámbito de la tesis, se han publicado en este foro.

Estos encuentros constituyen un buen punto de referencia para establecer las líneas básicas de investigación en el área de la combinación de clasificadores. Con anterioridad al año 2000, en el ámbito del reconocimiento de patrones, el aprendizaje automático y en algunas aplicaciones relativas a la fusión de datos se han establecido diversas pautas para la combinación. No obstante, a pesar de estos y otros muchos esfuerzos, todavía no

se ha definido una taxonomía general para la combinación de clasificadores, coexistiendo diferentes estructuras y planteamientos simultáneamente.

Siguiendo la filosofía establecida por Kuncheva (2004), la combinación de clasificadores puede verse desde distintas perspectivas o niveles, tal y como se muestra en la figura 3.7.

El nivel de datos, D, se refiere a la utilización de diferentes subconjuntos de datos para la clasificación. Desde la perspectiva de la clasificación de imágenes, podría hablarse de datos procedentes de imágenes distintas. El nivel C se refiere a la combinación de propiedades o atributos de las características a clasificar; ciñéndonos de nuevo al caso de las imágenes, los píxeles serían las características y como propiedades de las mismas podríamos utilizar las componentes de color, el módulo y dirección del gradiente u otras, que se combinarían en este nivel. El nivel B hace referencia al uso de distintos clasificadores individuales (Cl_1, Cl_2, \dots, Cl_L) utilizados para la combinación. Finalmente, el nivel A está relacionado con la estrategia utilizada para llevar a cabo la combinación.

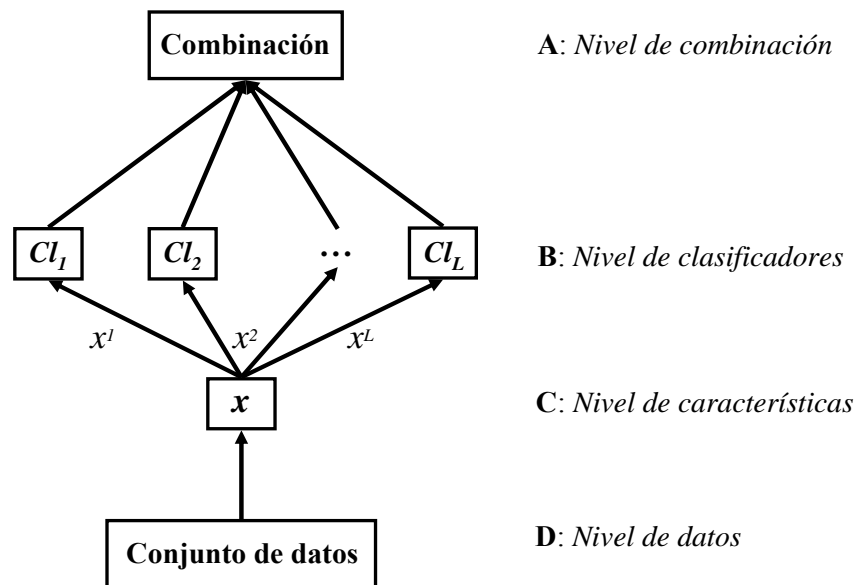


Figura 3.7 Niveles en la combinación de clasificadores

En el presente trabajo, la combinación se aborda desde los niveles A y B, principalmente desde el A. En efecto, es en el nivel A donde se han diseñado las estrategias de combinación, que se describen en el capítulo cuatro a partir de los clasificadores individuales descritos en la sección 3.3 de este mismo capítulo. Como se verá en el capítulo cuatro, dependiendo de la estrategia de combinación elegida, el número de clasificadores para la combinación L varía, siendo en unos casos seis y en otros dos. A partir de este momento, nos centramos exclusivamente en el nivel A.

3.5.2. Métodos para combinar las salidas de los clasificadores

Como se verá posteriormente, la combinación de clasificadores se puede realizar tanto en la fase de entrenamiento como en la de decisión. En la fase de decisión el patrón \mathbf{x}_s debe clasificarse como perteneciente a una clase w_j de acuerdo con dicha combinación. En adelante, por simplicidad y sin pérdida de generalidad se suprime el subíndice del vector representativo de la muestra \mathbf{x}_s identificándola como \mathbf{x} , de este modo se utiliza este vector genérico tanto para las fases de entrenamiento como de decisión.

Refiriéndonos a la figura 3.7 se dispone de un conjunto de L clasificadores en el nivel B, pertenecientes a la categoría de individuales, cuyas salidas deben combinarse en el nivel A. Por tanto, se entiende que desde esta perspectiva, la combinación se realiza en la fase de decisión.

No obstante, antes de abordar la problemática de la combinación distinguimos entre tres tipos de salidas, que pueden ser proporcionadas de forma general por los clasificadores (Xu y col., 1992; Kuncheva 2004):

Tipo 1 (nivel abstracto): cada clasificador Cl_i genera una etiqueta $s_i \in \Omega$ para la clase w_i , con $i = 1, 2, \dots, L$, de forma que para cada muestra a clasificar con vector de entrada \mathbf{x} , los L clasificadores definen un vector $\mathbf{s} = [s_1, s_2, \dots, s_L] \in \Omega^L$. En este nivel de abstracción no existe información sobre las etiquetas seleccionadas, ni existen etiquetas alternativas que puedan sugerirse. Por definición, cualquier clasificador es capaz de generar una etiqueta para \mathbf{x} . Las salidas de este tipo son las más generales de todas.

Tipo 2 (nivel de rango): la salida de cada clasificador Cl_i es un subconjunto de Ω , con las opciones de clase ordenadas según el grado de posibilidad de ser la etiqueta correcta (Ho y col., 1994; Tubbs y Alltop, 1991). Este tipo es deseable para problemas que requieran un número elevado de clases.

Tipo 3 (nivel de medida): cada clasificador Cl_i genera como salida un vector c -dimensional $[d_{i,1}, \dots, d_{i,j}, \dots, d_{i,c}]^T$, cuyas componentes son en realidad funciones del vector de entrada \mathbf{x} representativo de la muestra a clasificar. Si bien, para simplificar la notación se emplea $d_{i,j}$ en lugar de $d_{i,j}(\mathbf{x})$, que sería el correcto. El valor $d_{i,j}$ representa el grado de apoyo dado por el clasificador Cl_i para la muestra \mathbf{x} respecto de la clase w_j . A este tipo

pertenecen los clasificadores individuales estudiados en la sección 3.3. Como se ha mencionado previamente, en este trabajo de investigación, la combinación se realiza en el nivel A utilizando las salidas proporcionadas por los clasificadores que participan en la combinación; en la tabla 3.1 se sintetizan dichas salidas para la fase de decisión.

Por tanto, centrándonos en los clasificadores de tipo 3, a continuación se exponen diversos métodos de combinación clásicos para combinar las salidas de los clasificadores individuales. Posteriormente, en el capítulo cuatro se proponen las nuevas estrategias de combinación de las salidas, fruto de la investigación realizada. Los métodos combinados, descritos a continuación en este mismo capítulo, sirven para comparar y validar las estrategias combinadas propuestas en el capítulo cuatro.

3.5.2.1. Votación Mayoritaria

Se trata del famoso método de consenso, de ahí su nombre original en terminología inglesa de “*majority voting*”. La combinación se puede llevar a cabo mediante tres enfoques diferentes: unanimidad, mayoría simple y pluralidad. En el caso de la unanimidad, todos los clasificadores coinciden en su decisión, la mayoría simple se refiere a la coincidencia del 50% más uno y finalmente la pluralidad consiste en seleccionar la clase sobre la que coinciden como la correcta el mayor número de clasificadores.

Supóngase que las etiquetas de salida de los clasificadores vienen dadas como vectores c -dimensionales binarios $[d_{i,1}, d_{i,2}, \dots, d_{i,c}]^T \in \{0,1\}^c$, $i = 1, \dots, L$, donde $d_{i,j}$ es igual a 1 si el clasificador Cl_i etiqueta la muestra x como perteneciente a w_j y cero en caso contrario. El voto plural llega a ser una decisión de conjunto para la clase w_k si,

$$\sum_{i=1}^L d_{i,k} = \max_{j=1}^c \sum_{i=1}^L d_{i,j} \quad (3.45)$$

El voto plural se conoce en sentido amplio como votación mayoritaria, su planteamiento proviene del contexto electoral habiendo sido utilizado en diferentes trabajos de investigación, entre los que destacan Lin y col. (2003), Kittler y col. (1998), Lam y Suen (1997) o Battiti y Colla (1994).

Una de las variantes que se suelen introducir en este tipo de combinación es la del voto ponderado, en el sentido de que cada clasificador posee un peso específico que se tiene en cuenta a la hora de tomar la decisión (Kuncheva, 2004).

3.5.2.2. Combinación de Bayes

Se trata de otro de los esquemas de combinación de clasificadores clásico, donde se asume el hecho de que los clasificadores son mutuamente independientes entre sí. Esto es, a la hora de asignar una etiqueta de clase a la muestra definida por \mathbf{x} , los clasificadores toman esta decisión de forma independiente (independencia condicional). Esta es la razón por la que se le denomina *modelo independiente* (Titterington y col., 1981). En la literatura también aparece como *Bayes simple* o “*naive Bayes*” en terminología inglesa (Domingos y Pazzani, 1997). Si bien, en general, es conocido sencillamente como *combinación de Bayes* (Kuncheva, 2004).

Sea $P(s_i / w_j)$ la probabilidad con la que el clasificador Cl_i determina, a partir de la salida que produce, el grado de pertenencia de la muestra \mathbf{x} a la clase w_j . Por tanto, dada la clase w_j y teniendo en cuenta la independencia condicional, se obtiene la siguiente expresión,

$$P(s / w_j) = P(s_1, s_2, \dots, s_L / w_j) = \prod_{i=1}^L P(s_i / w_j) \quad (3.46)$$

Por consiguiente, la probabilidad a posteriori requerida para etiquetar \mathbf{x} es,

$$P(w_j / s) = \frac{P(w_j) P(s / w_j)}{P(s)} = \frac{P(w_j) \prod_{i=1}^L P(s_i / w_j)}{\sum_{k=1}^c P(s / w_k)}; \quad j = 1, \dots, c \quad (3.47)$$

El denominador en la ecuación (3.47) no depende de w_j y por tanto puede ignorarse a efectos de decisión; en este caso, el apoyo total resultante de la combinación recibido por la muestra \mathbf{x} para ser asignada a la clase w_j , es decir, la salida de la combinación resulta ser la dada en la ecuación (3.48).

$$\gamma_j(\mathbf{x}) = P(w_j) \prod_{i=1}^L P(s_i / w_j) \quad (3.48)$$

Como se verá en la sección siguiente, esta combinación puede asimilarse a la función producto, que se define allí.

3.5.2.3. Combinación mediante funciones

Siguiendo con los clasificadores de tipo 3, en esta sección se introducen una serie de funciones para combinar las salidas de los clasificadores.

Sin pérdida de generalidad, se supone que las salidas generadas por cada clasificador toman valores continuos en el intervalo $[0,1]$. Recordemos que $d_{i,j}(\mathbf{x})$ expresa el grado de apoyo dado por el clasificador Cl_i a la muestra \mathbf{x} según su pertenencia a la clase w_j . Cuanto mayor sea su valor para una determinada clase w_j , tanto más probable resulta el hecho de que dicha muestra pertenezca a esa clase. Volviendo sobre la tabla 3.1, donde se proporcionan las salidas de los seis clasificadores expresados allí, para los tres primeros el mayor grado de apoyo se corresponde con los valores máximos del grado de pertenencia y de las probabilidades; mientras que para los otros tres, al tratarse de distancias, es necesario realizar la transformación siguiente $\left[1 + d(\mathbf{x}, \mathbf{c}_j)\right]^{-1}$, de forma que si la distancia es cero, el grado de apoyo resulta ser la unidad y si la distancia es la máxima posible en el cubo representativo del modelo de color RGB, es decir la que corresponde a la diagonal principal que une los puntos $(0,0,0)$ y $(255, 255, 255)$, esto es $255\sqrt{3}$, el grado de apoyo se aproxima a cero.

Las salidas de los L clasificadores para una muestra dada se pueden organizar en forma de matriz, conocida como *perfil de decisión*, $PD(\mathbf{x})$,

$$PD(\mathbf{x}) = \begin{bmatrix} d_{1,1}(\mathbf{x}) & \cdots & d_{1,j}(\mathbf{x}) & \cdots & d_{1,c}(\mathbf{x}) \\ \vdots & & \vdots & & \vdots \\ d_{i,1}(\mathbf{x}) & \cdots & d_{i,j}(\mathbf{x}) & \cdots & d_{i,c}(\mathbf{x}) \\ \vdots & & \vdots & & \vdots \\ d_{L,1}(\mathbf{x}) & \cdots & d_{L,j}(\mathbf{x}) & \cdots & d_{L,c}(\mathbf{x}) \end{bmatrix} \quad (3.49)$$

Las filas se refieren a las salidas proporcionadas por cada clasificador Cl_i con respecto a las diferentes clases w_j , con $j = 1, \dots, c$; cada columna representa el apoyo proporcionado por los L clasificadores Cl_1, \dots, Cl_L para que \mathbf{x} pertenezca a la clase w_j . Los métodos descritos a continuación utilizan la matriz $PD(\mathbf{x})$ para encontrar la salida resultante de la combinación, que representa el apoyo total dado a \mathbf{x} por la combinación de clasificadores respecto a su pertenencia a una clase determinada. Existen dos enfoques para realizar esta tarea. En primer lugar se puede utilizar el hecho de que los valores en cada columna

j de la matriz PD representan los apoyos individuales que cada clasificador otorga a x para la clase w_j , la combinación de estos valores será el apoyo total proporcionado por los clasificadores combinados. La combinación de métodos que utilizan una columna de $PD(x)$ se denominan “class conscious” en Kuncheva y col. (2001). A esta categoría pertenecen la *media simple* y *promediada*, el *producto* y las estadísticas de *orden*, que son, como se verá seguidamente, algunos de los métodos empleados para realizar el análisis comparativo con los métodos combinados propuestos en este trabajo de investigación.

Alternativamente, se puede suprimir el concepto de clase y tratar los valores $d_{i,j}(x)$ como características en un nuevo espacio de características, denominado *espacio de características intermedio*. La decisión final se realiza mediante otro clasificador que toma este espacio de características intermedio como entrada y genera una etiqueta de clase a la salida. En Kuncheva y col. (2001) este tipo de métodos se denominan “class indifferent”; de esta forma se pueden construir diversas capas de clasificadores. El principal problema de este segundo tipo de clasificadores combinados estriba en cómo entrenar dichas arquitecturas para estar seguros de que el aumento de la complejidad mejore la eficacia de la combinación. Esto conlleva por otro lado, que la combinación se realice tanto en la fase de entrenamiento como de decisión, a diferencia de las estrategias propuestas en este trabajo en las que la combinación se realiza únicamente en la fase de decisión. Por las razones expresadas, este tipo de estrategias no se han considerado a efectos de establecer comparaciones de resultados entre ellas y las propuestas en el capítulo cuatro.

Centrándonos en la primera categoría, a continuación se describen una serie de clasificadores combinados pertenecientes a la misma. Kuncheva (2004) llama *no entrenables* a este tipo de clasificadores combinados, para expresar que la combinación no tiene parámetros extra que deban ser obtenidos mediante entrenamiento durante la fase de decisión.

El apoyo total, mencionado previamente, proporcionado por los L clasificadores, Cl_1, \dots, Cl_L , para que x pertenezca a la clase w_j es el resultado de la combinación de los valores en la j -ésima columna de $PD(x)$ según la expresión (3.50).

$$\gamma_j(\mathbf{x}) = \mathfrak{S}[d_{1,j}(\mathbf{x}), \dots, d_{i,j}(\mathbf{x}), \dots, d_{L,j}(\mathbf{x})] \quad (3.50)$$

donde \mathfrak{S} define la función de combinación. La decisión sobre la clase a la que pertenece \mathbf{x} se determina como el índice j correspondiente al máximo de los $\gamma_j(\mathbf{x})$ valores, en definitiva según la siguiente regla: $\mathbf{x} \in w_j$ si $\gamma_j(\mathbf{x}) > \gamma_k(\mathbf{x}) \forall k \neq j$, donde $j, k = 1, \dots, c$. La función \mathfrak{S} puede elegirse de varias maneras diferentes, las elecciones más comúnmente utilizadas vienen dadas por las ecuaciones (3.51) a (3.54) (Kittler y col., 1998).

Media aritmética

$$\gamma_j(\mathbf{x}) = \frac{1}{L} \sum_{i=1}^L d_{i,j}(\mathbf{x}) \quad (3.51)$$

Mínimo/Máximo

$$\gamma_j(\mathbf{x}) = \min_i \{d_{i,j}(\mathbf{x})\} \quad \gamma_j(\mathbf{x}) = \max_i \{d_{i,j}(\mathbf{x})\} \quad (3.52)$$

Producto

$$\gamma_j(\mathbf{x}) = \prod_{i=1}^L d_{i,j}(\mathbf{x}) \quad (3.53)$$

Media generalizada (Dubois y Prade, 1985)

$$\gamma_j(\mathbf{x}, \alpha) = \left(\frac{1}{L} \sum_{i=1}^L d_{i,j}(\mathbf{x})^\alpha \right)^{\frac{1}{\alpha}} \quad (3.54)$$

Tomando como referencia la media generalizada, si $\alpha = -1$, se tiene la media armónica y si $\alpha = 1$, lo que se obtiene es la media aritmética. La media armónica es muy sensible a valores pequeños de $d_{i,j}(\mathbf{x})$; pues bien, dado que en los experimentos llevados a cabo en este trabajo pueden aparecer valores de esta naturaleza, hemos optado por elegir la media aritmética definida en (3.51).

Por otro lado, la función producto puede considerarse como una generalización de la combinación de Bayes, descrita en la sección 3.5.2.2. En efecto, tomando como referencia la ecuación (3.48), en ella se combinan bajo el producto las probabilidades de pertenencia de \mathbf{x} a la clase w_j dadas por los L clasificadores, mientras que en la expresión (3.53) la combinación consiste en el producto de los valores $d_{i,j}(\mathbf{x})$, que pueden o no ser probabilidades. Por tanto, cuando son probabilidades estamos en el caso de la

combinación de Bayes, con la excepción de la probabilidad a priori $P(w_j)$, que aparece en la expresión (3.48) y que no existe en la (3.53). Esta probabilidad expresa el grado de certeza que se tiene sobre la pertenencia de x a la clase w_j con anterioridad al proceso de decisión, en base a algún tipo de conocimiento previo. Pues bien, dado que en nuestro caso carecemos de este conocimiento, resulta que una muestra x tiene la misma certeza a priori de pertenecer a cualquier clase y por tanto esta probabilidad no interviene en el proceso de decisión. En consecuencia, la función producto engloba bajo su definición a la combinación de Bayes. En resumen, para la comparación de los resultados obtenidos con los métodos de combinación propuestos en el capítulo cuatro, de los métodos anteriores se seleccionan los siguientes: votación mayoritaria, media aritmética, máximo, mínimo y producto.

3.5.2.4. Combinación mediante operadores de agregación fuzzy

Yager (1988) propuso la utilización de un operador de agregación en el ámbito de la teoría de la lógica fuzzy desde el punto de vista de la teoría de la decisión multi-criterio. Concretamente se trata de la *intersección de conjuntos borrosos*.

Siguiendo la teoría expuesta en Zimmermann (1991), un conjunto clásico se define normalmente como una colección de elementos $x \in X$ que puede ser finito, contable o incontable. Cada elemento puede o no pertenecer a un conjunto A , $A \subseteq X$. En el primer caso la sentencia “ x pertenece a A ” es cierta, mientras que en el segundo es falsa.

Por el contrario, se define un conjunto fuzzy \tilde{A} en X como un conjunto de pares ordenados $\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) / x \in X\}$; $\mu_{\tilde{A}}(x)$ se conoce como grado de pertenencia de x en \tilde{A} , que toma valores en el rango $[0,1]$. Dados dos conjuntos fuzzy \tilde{A} y \tilde{B} , la intersección se define como sigue,

$$\tilde{A} \cap \tilde{B} = \{(x, \mu_{\tilde{A} \cap \tilde{B}}(x)) / x \in X\} \quad (3.55)$$

donde

$$\mu_{\tilde{A} \cap \tilde{B}}(x) = 1 - \min \left\{ 1, \left((1 - \mu_{\tilde{A}}(x))^p + (1 - \mu_{\tilde{B}}(x))^p \right)^{\frac{1}{p}} \right\} \quad \text{con } p \geq 1 \quad (3.56)$$

Realizando un proceso de abstracción y asumiendo que los elementos de la matriz PD , definida en (3.49) pueden considerarse grados de pertenencia fuzzy, si se restringen sus valores al rango $[0, +1]$, entonces es posible combinar dichos valores mediante el operador de agregación de intersección definido previamente. Suponiendo que solamente tuviésemos dos clasificadores (Cl_h, Cl_k) , y teniendo en cuenta la matriz PD combinaríamos cada una de sus columnas mediante la ecuación (3.56) como sigue,

$$\gamma_j^{hk}(\mathbf{x}) = 1 - \min \left\{ 1, \left((1 - d_{h,j}(\mathbf{x}))^p + (1 - d_{k,j}(\mathbf{x}))^p \right)^{\frac{1}{p}} \right\} \quad \text{con } p \geq 1 \quad (3.57)$$

donde p es un parámetro que debe fijarse previamente, lo cual se realiza mediante el proceso descrito en el capítulo cinco, sección 5.3.2.2.

Pues bien, dado que la intersección de conjuntos es asociativa, la extensión de la anterior formulación, cuando en la combinación intervienen más de dos clasificadores, se realiza tomando dos a dos los valores de cada columna en la matriz PD y combinando nuevamente de dos en dos los valores resultantes. La decisión final se toma de acuerdo con la ya conocida regla: $\mathbf{x} \in w_j$ si $\gamma_j(\mathbf{x}) > \gamma_k(\mathbf{x}) \quad \forall k \neq j$, donde $j, k = 1, \dots, c$.

Bajo la filosofía de la combinación de los operadores de agregación fuzzy siempre es posible utilizar diferentes definiciones de agregación tanto desde el punto de vista de la intersección como de la unión de conjuntos fuzzy, tales como Hamacher, Dubois y Prade, Werners o Zimmermann y Zysno (Zimmermann, 1991). No obstante, en el presente trabajo no se ha estimado conveniente incluir más operadores de agregación ya que diversas pruebas experimentales realizadas con ellos no han aportado ninguna mejora respecto de la intersección de Yager (1998), prefiriendo por ello utilizar ésta para la comparación de los resultados obtenidos por los métodos propuestos en el capítulo cuatro.

3.5.2.5. Resumen de los modos de combinación de clasificadores

En la figura 3.8 se muestra un esquema gráfico sobre el mecanismo de combinación de los clasificadores descritos en las dos secciones previas. Este esquema está basado en la idea expuesta en Kuncheva (2004). Por otro lado, en la tabla 3.2 se resumen las funciones de combinación de esos mismos clasificadores según las definiciones previas, así como las reglas para la toma de decisiones sobre la asignación de \mathbf{x} a una determinada clase w_j .

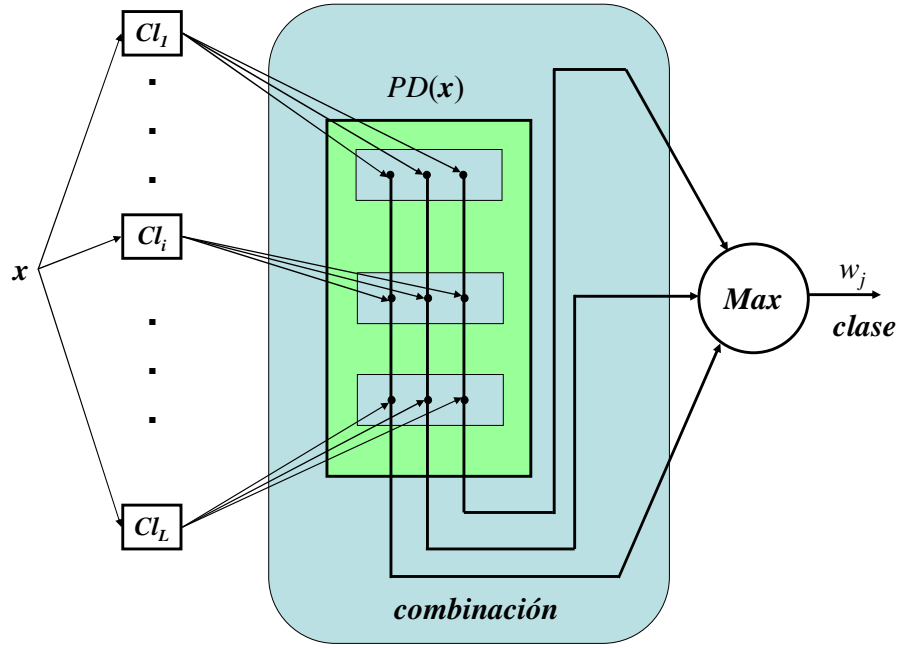


Figura 3.8 Combinación de clasificadores individuales

Tabla 3.2 Resumen de métodos de combinación y reglas de decisión

	Clase 1 (w_1)	Clase 2 (w_2)		Clase c (w_c)
Clasificador Cl_1	$d_{1,1}(x)$	$d_{1,2}(x)$...	$d_{1,c}(x)$
Clasificador Cl_2	$d_{2,1}(x)$	$d_{2,2}(x)$...	$d_{2,c}(x)$
\vdots	\vdots	\vdots		\vdots
Clasificador Cl_L	$d_{L,1}(x)$	$d_{L,2}(x)$...	$d_{L,c}(x)$
<i>Media aritmética</i>	$\gamma_1(x) = \frac{1}{L} \sum_{i=1}^L d_{i,1}(x)$	$\gamma_1(x) = \frac{1}{L} \sum_{i=1}^L d_{i,2}(x)$...	$\gamma_c(x) = \frac{1}{L} \sum_{i=1}^L d_{i,c}(x)$
<i>Máximo</i>	$\gamma_1(x) = \max_1 \{d_{i,1}(x)\}$	$\gamma_2(x) = \max_2 \{d_{i,2}(x)\}$...	$\gamma_c(x) = \max_c \{d_{i,c}(x)\}$
<i>Mínimo</i>	$\gamma_1(x) = \min_1 \{d_{i,1}(x)\}$	$\gamma_2(x) = \min_2 \{d_{i,2}(x)\}$...	$\gamma_c(x) = \min_c \{d_{i,c}(x)\}$
<i>Producto</i>	$\gamma_1(x) = \prod_{i=1}^L d_{i,1}(x)$	$\gamma_2(x) = \prod_{i=1}^L d_{i,2}(x)$...	$\gamma_c(x) = \prod_{i=1}^L d_{i,c}(x)$
<i>Intersección de Yager</i>	$\gamma_1(x) = \bigcap_{hk} \gamma_1^{hk}(x)$	$\gamma_2(x) = \bigcap_{hk} \gamma_2^{hk}(x)$...	$\gamma_c(x) = \bigcap_{hk} \gamma_c^{hk}(x)$
Decisión final $\max_{j=1}^c \{\gamma_j(x)\}$				

3.5.2.6. Combinación mediante la integral fuzzy

La integral fuzzy se ha utilizado en diferentes aplicaciones con resultados prometedores actuando como combinación de clasificadores (Kuncheva, 2003; Verikas y Lipnickas, 2002; Verikas y col., 1999; Cho y Kim, 1995 *a,b*). Obviando los fundamentos teóricos que la sustentan, a continuación se presenta el algoritmo utilizado para la clasificación desde el punto de vista de la combinación de clasificadores.

El principal problema con el uso de la integral fuzzy, como estrategia de combinación, estriba en la necesidad de conocer lo que se denomina la densidad fuzzy g para cada clasificador individual. Estas densidades, en cierto modo, representan las competencias relativas g^i de cada clasificador individual Cl_i , según algún criterio preestablecido. Afortunadamente, en el caso que nos ocupa disponemos ya de un mecanismo para obtener esas competencias por medio de la ecuación (3.44).

Por otro lado, para aplicar el algoritmo de la integral fuzzy es necesario conocer la denominada medida λ - fuzzy utilizando los datos. El valor de λ se obtiene como la única raíz real mayor que -1 del polinomio siguiente,

$$\lambda + 1 = \prod_{i=1}^L (1 + \lambda g^i) \quad (3.58)$$

Obtenidas las competencias relativas de cada clasificador g^1, g^2, \dots, g^L y calculada λ , la operación de la integral fuzzy funcionando como un clasificador combinado se sintetiza como sigue:

1. Dada la muestra \mathbf{x} a clasificar, se ordena la k -ésima columna de la matriz $PD(\mathbf{x})$ en orden decreciente obteniendo $[d_{i_1,k}(\mathbf{x}), d_{i_2,k}(\mathbf{x}), \dots, d_{i_L,k}(\mathbf{x})]^T$, siendo $d_{i_1,k}(\mathbf{x})$ el máximo valor y $d_{i_L,k}(\mathbf{x})$ el menor. T expresa transposición.
2. Ordenar las densidades fuzzy o competencias relativas convenientemente y por tanto en el siguiente orden g^{i_1}, \dots, g^{i_L} y hacer $g(1) = g^{i_1}$
3. Desde $t = 2$ hasta L , calcular recursivamente,

$$g(t) = g^{i_t} + g(t-1) + \lambda g^{i_t} g(t-1) \quad (3.59)$$

4. Obtener el valor de pertenencia de la muestra x a la clase w_j mediante,

$$\gamma_j(x) = \max_{t=1}^L \left\{ \min \left\{ d_{i,j}(x), g(t) \right\} \right\} \quad (3.60)$$

El valor de pertenencia final $\gamma_j(x)$ para la clase w_j , representa el compromiso entre las *competencias* (expresadas por las densidades fuzzy g) y la *evidencia* representada por la columna j -ésima de la matriz $PD(x)$. Obsérvese que el vector de medidas fuzzy $[g(1), \dots, g(L)]^T$ podría ser diferente para cada clase e incluso para cada muestra x . No obstante, en los experimentos llevados a cabo en este trabajo de investigación, estas dos posibilidades no se han considerado.

La decisión final se toma a partir de la conocida regla de decisión dada por,

$$x \in w_j \text{ si } \gamma_j(x) > \gamma_k(x) \quad \forall k/k \neq j \quad (3.61)$$

La figura 3.9 representa el modo de operación del método de combinación de clasificadores basado en la integral fuzzy.

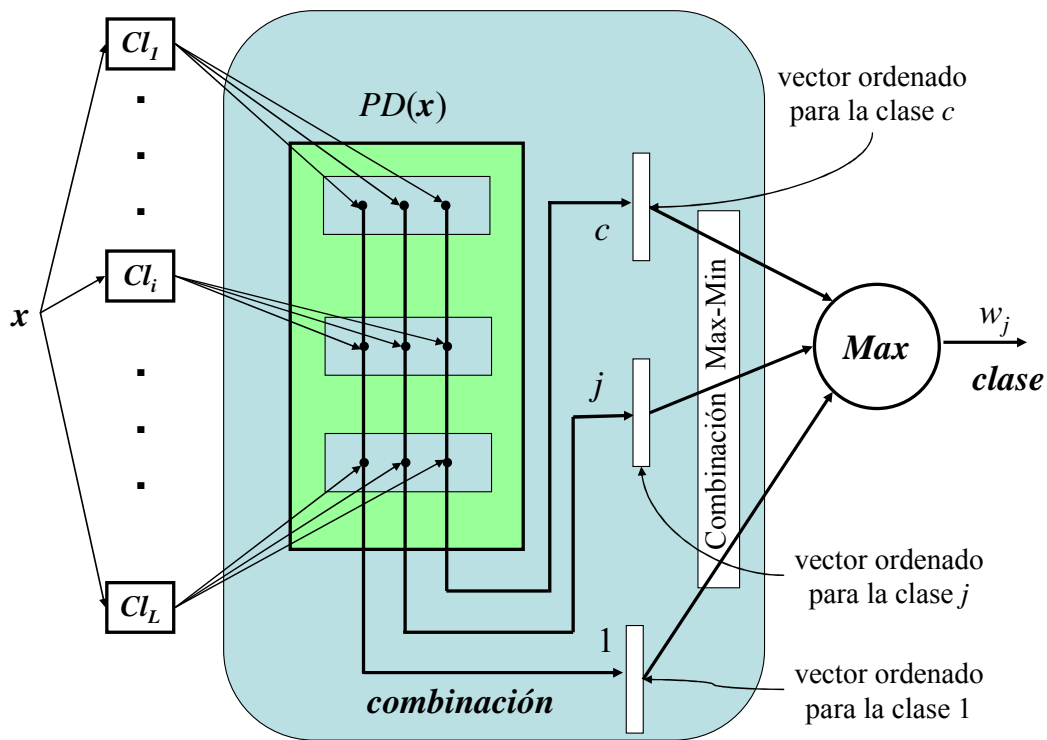


Figura 3.9 Esquema de funcionamiento de la integral fuzzy en la combinación de clasificadores

Esta figura tiene una más fácil interpretación si se toma como referencia la tabla 3.3, donde cada clasificador tiene asociada su competencia relativa en la columna correspondiente. El procedimiento para obtener las competencias relativas se pospone al capítulo cuatro, sección 4.2.3. Los elementos de cada columna en la matriz PD , es decir de cada clase, se ordenan de mayor a menor para su posterior combinación con las competencias según el procedimiento descrito previamente, procediendo finalmente a tomar la decisión final según la regla dada en (3.60).

Tabla 3.3 Competencias relativas de cada clasificador

	Clase 1 (w_1)	Clase 2 (w_2)		Clase c (w_c)	Competencias relativas
Clasificador Cl_1	$d_{1,1}(x)$	$d_{1,2}(x)$...	$d_{1,c}(x)$	g^1
Clasificador Cl_2	$d_{2,1}(x)$	$d_{2,2}(x)$...	$d_{2,c}(x)$	g^2
\vdots	\vdots	\vdots		\vdots	
Clasificador Cl_L	$d_{L,1}(x)$	$d_{L,2}(x)$...	$d_{L,c}(x)$	g^L

El algoritmo descrito anteriormente calcula la denominada *integral fuzzy de Sugeno*. El segundo tipo más popular de la integral fuzzy es la denominada *integral fuzzy de Choquet* (Grabisch, 1995), que utilizan las mismas medidas λ – fuzzy ; la única diferencia estriba en que la ecuación (3.60) debe reemplazarse por la siguiente,

$$\gamma_j(x) = d_{i,j}(x) + \sum_{k=2}^L [d_{i_{k-1},j}(x) - d_{i_k,j}(x)] g(k-1) \quad (3.62)$$

En los experimentos llevados a cabo con esta expresión se han obtenido resultados muy similares a los anteriores, por lo que a partir de este momento todas las referencias a la integral fuzzy, se referirán a los cálculos realizados con la ecuación (3.60).

En cualquier caso, el principal inconveniente de ambas integrales estriba en la necesidad del cálculo de las competencias relativas de los diferentes clasificadores. Este problema ha sido resuelto satisfactoriamente gracias a la obtención de las competencias relativas tal y como se ha planteado previamente.

3.6. Resumen del capítulo

En este capítulo se han descrito los seis clasificadores individuales, a saber: agrupamiento borroso, estimador paramétrico de máxima verosimilitud, estimador no paramétrico de Parzen, mapas auto-organizativos y cuantización vectorial en sus dos versiones. Estos clasificadores constituyen la base de las estrategias combinadas propuestas en el trabajo de investigación que se describen en el capítulo cuatro.

Se han estudiado básicamente tres funciones criterio (PC , SE y VC) asociadas con dichos clasificadores, que permiten validar las particiones de datos disponibles inicialmente, consiguiendo de este modo el diseño de las estrategias no supervisadas descritas en el capítulo cuatro. De este modo, se logra un procedimiento de clasificación automático. Tras los estudios realizados sobre las tres funciones criterio anteriores, finalmente se ha optado por utilizar solamente SE y VC . Se ha definido un procedimiento utilizando estas dos últimas funciones criterio mediante el cual es posible determinar las competencias relativas de los seis clasificadores individuales, gracias a lo cual, dichas competencias pueden utilizarse convenientemente durante las estrategias combinadas propuestas, como se describe posteriormente en el capítulo cuatro.

Además, en el capítulo se han introducido los conceptos elementales relativos a la combinación de clasificadores. Se ha optado por una serie de estrategias de combinación, por las razones expuestas, como las más apropiadas para comparar los resultados obtenidos por los clasificadores combinados propuestos en este trabajo. Los clasificadores finalmente seleccionados con tal propósito son: votación mayoritaria, media aritmética, máximo, mínimo, producto y operador de agregación de Yager.

Tanto para los clasificadores individuales como para los combinados se han definido las dos fases típicas de entrenamiento y decisión. Para cada una de dichas fases se han proporcionado los detalles suficientes para comprender los procesos involucrados. Se ha incidido en el hecho de que tanto las estrategias combinadas estudiadas en este capítulo como las propuestas en el capítulo cuatro, llevan a cabo la combinación durante la fase de decisión.

Capítulo 4

Estrategias locales y globales para la combinación de clasificadores individuales

4.1. Introducción

En el capítulo tres se han descrito seis métodos de clasificación catalogados como individuales, que constituyen la base de los métodos combinados propuestos en este capítulo. Dichos clasificadores son los que se indican a continuación, identificados a partir de ahora con la simbología que se indica: *a*) el método de Agrupamiento Borroso o “Fuzzy Clustering” (*FC*), *b*) el estimador paramétrico de máxima verosimilitud (*BP*), *c*) el estimador no paramétrico basado en la ventana de Parzen (*BN*), *d*) el método basado en los Mapas Auto-Organizativos o “Self-Organizing Maps” (*SO*) y *e*) dos versiones del método de cuantización vectorial “Learning Vector Quantization” (*L1* y *L2*). Se han descrito también varios métodos combinados de clasificadores resumidos en la tabla 3.2, con el fin de establecer un análisis comparativo respecto de las estrategias combinadas, propuestas en el presente capítulo. Tanto los métodos individuales como los combinados se han abordado desde las dos fases típicas involucradas en todo proceso de aprendizaje/clasificación, a saber: *entrenamiento* y *decisión*, figura 3.1.

En Duda y col. (2001) se plantea el diseño de un clasificador *no supervisado* a partir de otro *supervisado* durante la fase de entrenamiento. Básicamente, el procedimiento consiste en un proceso iterativo que comienza suponiendo la existencia de un cierto número de clases, de suerte que dado un clasificador, éste estima sus parámetros bajo dicha suposición. Posteriormente, se evalúa la calidad de la partición utilizando medidas

tales como las definidas en las ecuaciones (3.38) a (3.41). Si las medidas realizadas son satisfactorias según el criterio o criterios establecidos, la partición se considera válida. Este es el esquema seguido en el diseño de los métodos combinados no supervisados propuestos en este capítulo, como se verá posteriormente.

Además, en este mismo capítulo se presentan, como aporte fundamental del trabajo de investigación realizado, dos estrategias para la combinación de clasificadores en el contexto de las imágenes naturales de exterior. La primera realiza la combinación a nivel de píxel, catalogada por este motivo como de naturaleza *local*. La segunda considera para la combinación un píxel y sus vecinos, debido al uso de la vecindad, se le considera de carácter *global*. Las características en ambas son los píxeles, tanto si éstos intervienen aisladamente o formando parte de la vecindad. Las tres componentes espectrales de estas características en el modelo de color RGB, son las propiedades utilizadas, tal y como se ha descrito en el capítulo tres, sección 3.2.1. En ambos enfoques, la combinación de los clasificadores se realiza durante la fase de decisión.

Bajo la cobertura del enfoque local se proponen a su vez dos métodos, que utilizan los seis clasificadores individuales citados previamente, esto es: *FC*, *BP*, *BN*, *SO*, *L1* y *L2*. El primer método se basa en el paradigma de la *teoría de la Toma de Decisiones Multi-Criterio*, conocida en la literatura científica internacional especializada como “*Multi-Criteria Decision Making, MCDM*”, bajo la perspectiva de la lógica *Fuzzy*. A este método nos referiremos de aquí en adelante con las siglas bajo la nomenclatura internacional *MCDMF*, por su amplia difusión. La propuesta formulada en este capítulo tiene su origen en el trabajo de Guijarro y Pajares (2008). El segundo método de naturaleza local se fundamenta en la teoría de la *Integral Fuzzy (IF)* cuya base teórica se encuentra en Kuncheva (2004).

Como bien es sabido, las texturas en las imágenes no se configuran solamente con píxeles aislados, antes bien, se disponen de forma que entre píxeles vecinos se establecen ciertas relaciones estructurales. Este es el fundamento sobre el que se sustenta la segunda estrategia de naturaleza global. Antes de concretar los aspectos relacionados con esta estrategia vamos a sentar sus bases y fijar una serie de conceptos previos. Con tal propósito, podemos recurrir por ejemplo, a la imagen de la figura 3.5; en ella se observan diferentes texturas de forma que las propiedades espectrales de los píxeles que constituyen dichas texturas son similares. Más concretamente, si nos

fijamos por ejemplo en un píxel con tonalidades verdes, éste está normalmente rodeado de píxeles con tonalidades similares, excepción hecha de los píxeles frontera que delimitan dos o más texturas diferentes. Esta consideración, llevada al ámbito de la clasificación, se traduce en que dicho píxel junto con sus vecinos deberían identificarse probablemente como pertenecientes a la misma clase o equivalentemente, que los valores de salida proporcionados por los clasificadores individuales en la fase de decisión, para todos esos píxeles, deberían ser similares, salvo en el caso de píxeles fronterizos por las razones expuestas. Para centrar un poco más las ideas anteriores recurrimos al trabajo de Wang (2005), el cual en el ámbito de las redes neuronales establece los principios sobre cómo los elementos sensores en general perciben los objetos en la escena. Dicho de otra manera, en él se realiza un estudio pormenorizado sobre cómo abordar el problema del análisis de una escena tridimensional desde la perspectiva de una imagen bidimensional. Pues bien, tanto en Wang (2005) como en Botte-Lecocq y col. (2007) se mencionan algunos de los grandes principios sobre agrupamientos de la escena en una imagen bidimensional. De este modo, para tratar las escenas del mundo real es necesario establecer algún tipo de similitud entre los elementos involucrados en dicha escena. La pregunta es: ¿qué determina la relación de similitud entre los elementos de la escena? En el dominio visual, estas cuestiones se vienen estudiando desde hace mucho tiempo bajo la influencia de algunas corrientes psicológicas como las establecidas por la escuela de la Gestalt (Koffka, 1935). A continuación se proporciona una lista de principios aplicables a las imágenes bidimensionales bajo la perspectiva de esta escuela, según Palmer (1999):

- *Proximidad*: los píxeles que se sitúan próximos en el espacio tienden a agruparse entre sí.
- *Similitud*: los elementos que tienen propiedades similares, tales como luminancia, color, textura, etc., tienden igualmente a agruparse entre ellos.
- *Destino común*: los elementos que se mueven coherentemente, es decir con un movimiento común, también tienen tendencia a constituirse en grupos. Esto puede verse como un caso particular de la similitud. No obstante se expresa separadamente con el fin de poner el énfasis en aspectos dinámicos de la escena analizada.
- *Continuidad uniforme*: un conjunto de elementos sin discontinuidades significativas tienden a aglutinarse.

- *Conectividad y región común*: los elementos conectados espacialmente forman agrupaciones, que definen ciertas regiones de la imagen.
- *Familiaridad*: un conjunto de elementos que pertenecen a la misma estructura de patrones tienden a formar grupos homogéneos.

La característica común derivada de los principios anteriormente expuestos, es la tendencia de los elementos en la escena a constituirse en grupos. En el caso que nos ocupa, relativo a las texturas naturales en las imágenes de exterior, estos elementos se concretan en las relaciones que se establecen entre un píxel y sus vecinos según sus propiedades espectrales.

Pues bien, bajo esta perspectiva se propone la estrategia de combinación global, en este caso utilizando dos clasificadores: *FC* y *BP*. Esta estrategia, por tanto, tiene en cuenta tanto el píxel sobre el que se va a tomar la decisión como sus vecinos, donde los principios de *proximidad*, *similitud*, *continuidad uniforme* y *conectividad* cobran especial relevancia. Bajo esta estrategia global se proponen a su vez dos métodos. El primero de ellos tiene su fundamento en el paradigma de los denominados *Mapas Cognitivos Fuzzy (MCF)*, introducidos y estudiados inicialmente por Kosko (1992). Con la misma filosofía y esquema de funcionamiento se propone un segundo método de combinación cuyos fundamentos se centran en los esquemas de *relajación (REL)*. Los planteamientos iniciales sobre los procesos de esta naturaleza se deben a Rosenfeld y col. (1976) o Hummel y Zucker (1983). Esta propuesta proviene de los trabajos realizados en Guijarro y col. (2008) al amparo de esta investigación.

Resumiendo, en el capítulo se proponen dos estrategias para la combinación de clasificadores durante la fase de decisión, una local a nivel de píxel bajo la cual se diseñan los métodos *MCDMF* e *IF* y otra global teniendo en cuenta las vecindades de los píxeles, que permite diseñar otros dos métodos, *MCF* y *REL*. A continuación se describen, por este orden, dichos métodos combinados, dando detalles de los mismos tanto en lo que respecta a la fase de entrenamiento como de decisión.

4.2 Diseño de un clasificador local no supervisado: MCDMF

La figura 4.1 muestra el esquema de diseño del clasificador no supervisado basado en el paradigma *MCDMF*. Como puede observarse, el procedimiento trabaja en las ya

mencionadas dos fases de entrenamiento y decisión. En esta sección se expone brevemente el mecanismo de funcionamiento durante ambas fases, poniendo el énfasis en la de decisión, ya que es en ella donde realmente se produce el aporte fundamental de este trabajo de investigación en relación a este clasificador combinado. Con tal propósito, en primer lugar se describe el proceso de entrenamiento donde intervienen los seis clasificadores individuales, que estiman los parámetros involucrados en los respectivos procesos. Posteriormente, se exponen las líneas generales del método *MCDMF* sobre el que se sustenta el procedimiento de combinación propuesto. Finalmente, se concreta la aplicación del método al problema de la clasificación de las texturas naturales, que constituye el objeto de la investigación.

La explicación de algunos detalles mostrados en la figura 4.1 tales como los relativos a criterios de beneficio y coste se posponen para su posterior explicación en las siguientes secciones.

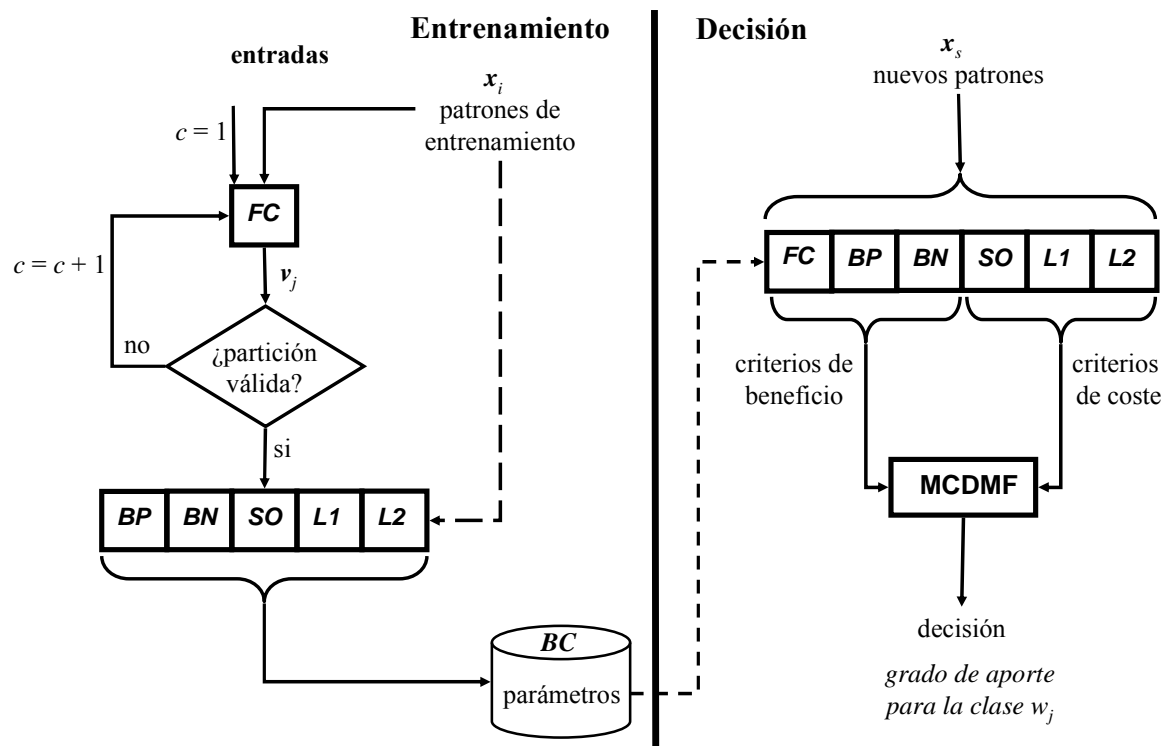


Figura 4.1 Diseño de un clasificador no supervisado combinado: fases de entrenamiento y decisión

4.2.1 Fase de entrenamiento

El proceso de entrenamiento comienza con el procesamiento de los patrones o muestras de entrenamiento disponibles en cada momento, constituyendo las entradas al sistema.

Inicialmente, se supone la existencia de una única clase y que todas las muestras pertenecen a dicha clase, por ello se hace $c = 1$. Bajo este supuesto, se desencadena el procedimiento de entrenamiento siguiendo el proceso descrito en la sección 3.3.1.1; llegándose a establecer una partición de las muestras en la única clase existente en este momento. La partición con una única clase se considera por definición no válida, ya que en las imágenes disponibles esta situación no se produce nunca. Por esta razón, se intenta una nueva partición con $c = 2$. Tras lo cual, se valora si la nueva partición es o no válida.

El proceso de validación se lleva a cabo mediante el cómputo de la variación relativa de los coeficientes SE y VC según la ecuación (3.43). El criterio de validación se establece teniendo en cuenta que la partición se da como válida cuando la variación relativa de dichos coeficientes, entre dos clases consecutivas, se sitúa por debajo de un umbral dado. La ecuación (3.43) permite calcular tanto $\hat{\Delta}^{SE}$ como $\hat{\Delta}^{VC}$ o $\hat{\Delta}^{PC}$, no obstante, en la sección 3.4.4 se exponen las razones por las que se descarta $\hat{\Delta}^{PC}$. Posteriormente, en la sección 5.3.1.5 se lleva a cabo un exhaustivo estudio por el que se determina que sea $\hat{\Delta}^{SE}$ el único criterio utilizado para validar la partición.

Si la partición no es válida, según el criterio anterior, el número de clases c se incrementa en una unidad, procediéndose de nuevo a repetir el proceso previo hasta conseguir la validación de la partición. La distribución de las muestras en las clases, una vez conocido el número de éstas, se lleva a cabo mediante el proceso pseudo-aleatorio descrito en la sección 3.3.1.1 según la teoría expuesta en Balasko y col. (2006). Este es el proceso básico que confiere a este diseño su naturaleza no supervisada según la filosofía establecida en Duda y col. (2001), como se ha mencionado previamente.

Por el contrario, cuando la partición es considerada válida, también según el mismo criterio anterior, los seis clasificadores individuales realizan sus correspondientes procesos de entrenamiento para llevar a cabo la estimación de sus parámetros. Estos procesos de entrenamiento son los descritos en las secciones del capítulo tres, que se indican a continuación: FC en 3.3.1.1, BP en 3.3.2.1, BN en 3.3.3.1, SO en 3.3.4.1 y finalmente $L1$ y $L2$ en 3.3.5.1.

Tras los respectivos procesos de entrenamiento, los parámetros estimados o aprendidos por cada uno de esos clasificadores son almacenados en la *Base de Conocimiento (BC)*,

quedando disponibles para su posterior recuperación y utilización durante la fase de decisión donde se produce la combinación de los clasificadores, figuras 3.1 y 4.1.

4.2.2 Descripción general del paradigma MCDM fuzzy

4.2.2.1 Formulación del problema y definiciones previas

La toma de decisiones es el proceso por el cual se intenta encontrar la mejor opción, ante un conjunto de alternativas posibles, según una serie de criterios disponibles. En la mayoría de los problemas de decisión se plantea la cuestión de resolver un problema con múltiples criterios. Un problema de esta naturaleza puede expresarse en forma matricial según se expresa en la ecuación (4.1). Para ello se dispone de o alternativas, $A_j (j=1,2,\dots,o)$ y n criterios $CR_i (i=1,2,\dots,n)$. Las alternativas determinan las opciones, que el clasificador puede elegir a la hora de tomar la decisión. Los criterios representan la base sobre los que dicha decisión va a ser tomada. Cada criterio CR_i , tiene asociado un peso específico z_i que determina la importancia o valor concedido al mismo. Con ello se definen las dos siguientes estructuras (Wang y Fenton, 2006): a) D , conocida como la matriz de decisión, en ella cada entrada x_{ji} representa la valoración de la alternativa A_j con respecto al criterio CR_i y b) Z , que representa el vector de pesos específicos asociados a cada criterio.

$$D = \begin{matrix} & \begin{matrix} CR_1 & CR_2 & \dots & CR_n \end{matrix} \\ \begin{matrix} A_1 \\ A_2 \\ \vdots \\ A_o \end{matrix} & \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{o1} & x_{o2} & \dots & x_{on} \end{bmatrix} \end{matrix} \text{ y } Z = [z_1, z_2, \dots, z_n], \text{ donde } i=1,2,\dots,n \text{ y } j=1,2,\dots,o \quad (4.1)$$

En forma abreviada, la ecuación (4.1) puede expresarse como sigue,

$$D = [x_{ji}]_{o \times n} \text{ y } Z = [z_i]_{1 \times n}, \text{ donde } i=1,2,\dots,n \text{ y } j=1,2,\dots,o \quad (4.2)$$

Los criterios se enmarcan según las dos categorías siguientes:

- *criterio de beneficio*: donde el valor más alto de x_{ji} representa la mejor opción a la hora de tomar la decisión con respecto a dicho criterio.

- *criterio de coste*: donde el valor más bajo de x_{ji} representa la mejor opción para la toma de la decisión.

En los métodos clásicos sobre toma de decisiones, los valores numéricos tanto de los criterios como de los pesos se conocen de forma precisa. Una revisión exhaustiva sobre tales métodos se puede encontrar en los trabajos de Hwang y Yoon (1981). Una de las primeras propuestas formales sobre la toma de decisiones se encuentra en dicho trabajo, siendo conocida como TOPSIS (“Technique for Order Performance by Similarity to Ideal Solution”), que basa la elección según el concepto de que la alternativa elegida será la que obtenga la mínima distancia a una solución ideal positiva, a la vez que se obtiene la máxima distancia a una solución ideal negativa.

Bajo ciertas condiciones, los datos precisos son inapropiados en algunos problemas de la vida real. Esto es debido a que los juicios emitidos, desde el punto de vista humano, sobre determinadas preferencias son a menudo vagos o imprecisos, no resultando fácil estimar su preferencia mediante un valor numérico exacto.

Una posible solución al modelo de valores exactos, proviene del ámbito de la lógica fuzzy. Tal es el caso de los primeros trabajos propuestos en Zadeh (1975), que utilizan etiquetas lingüísticas. Teniendo en cuenta este enfoque, tanto los criterios como las alternativas toman valores con un cierto grado de incertidumbre o borrosidad. Bajo esta perspectiva se redefine la ecuación (4.2) como sigue,

$$\tilde{D} = [\tilde{x}_{ji}]_{o \times n} \text{ y } \tilde{Z} = [\tilde{z}_i]_{1 \times n}, \text{ donde } i=1,2,\dots,n \text{ y } j=1,2,\dots,o \quad (4.3)$$

donde ahora \tilde{x}_{ji} representa la valoración fuzzy de la alternativa A_j con respecto al criterio CR_i ; y \tilde{z}_i determina el peso específico o importancia relativa fuzzy concedida al correspondiente criterio CR_i . En particular, un enfoque fácil y eficaz para capturar la incertidumbre por parte de un experto sobre el valor de un número desconocido, se puede obtener como un número triangular fuzzy. A partir de estos conceptos se definen del mismo modo las soluciones ideales tanto positiva como negativa fuzzy, por tanto, cada uno de los valores \tilde{x}_{ji} y \tilde{z}_j definidos en sendas matrices de la ecuación (4.3) y representando etiquetas lingüísticas, pueden describirse por números triangulares fuzzy.

Definición: un número triangular fuzzy \tilde{a} se puede definir por una tupla (a_1, a_2, a_3) . La función de pertenencia de un valor x dado, se define por medio de la ecuación (4.4) según Kaufmann y Gupta (1985),

$$\mu_{\tilde{a}}(x) = \begin{cases} (x - a_1)/(a_2 - a_1), & a_1 \leq x \leq a_2 \\ (a_3 - x)/(a_3 - a_2), & a_2 \leq x \leq a_3 \\ 0, & \text{en cualquier otro caso} \end{cases} \quad (4.4)$$

Los números triangulares fuzzy están basados en la evaluación de tres valores: el mínimo/máximo valor posible a_1/a_3 el valor más probable a_2 y el máximo/mínimo valor posible a_3/a_1 . En Chen (2000) se puede encontrar un estudio pormenorizado sobre las definiciones y propiedades de estos números fuzzy y las variables lingüísticas asociadas. De este modo, las variables lingüísticas que aparecen en la ecuación (4.3) pueden describirse específicamente mediante números triangulares fuzzy de la forma, $\tilde{x}_{ji} = (x_{ji1}, x_{ji2}, x_{ji3})$ y $\tilde{z}_i = (z_{i1}, z_{i2}, z_{i3})$.

4.2.2.2 Normalización

Los valores de los criterios pueden presentarse según diferentes escalas de valores. Esto es lo que ocurre en nuestra aplicación dado que las decisiones se basan en las salidas proporcionadas por los clasificadores, que representarán los criterios de decisión. Dichas salidas son por un lado grados de pertenencia (FC) o probabilidades (BP , BN) que varían en el rango $[0,1]$ y por otro distancias (SO , $L1$, $L2$), donde ahora los rangos de valores difieren de los anteriores. Por este motivo, para que durante la toma de decisiones el hecho de disponer de datos en diferentes rangos de valores no suponga un obstáculo, es necesario proceder a realizar un proceso de normalización.

Específicamente, se trata de normalizar los números fuzzy de la matriz de decisión, de forma que a partir de esta normalización se genere lo que denominamos matriz de decisión normalizada, también conocida en la terminología inglesa como “performance matrix”. Dicha matriz queda identificada en función de sus elementos como sigue (Wang y Fenton, 2006; Gu y Zhu, 2006),

$$\tilde{P} = [\tilde{p}_{ji}]_{o \times n} \quad (4.5)$$

Para evitar fórmulas complicadas de normalización, tal como las utilizadas en TOPSIS por Chen (2000), se propone una transformación de escalado lineal. Esta es la adoptada en este trabajo de investigación, la cual permite transformar los diferentes valores de los criterios en valores comparables entre sí. Dicha transformación se concreta como sigue,

$$\tilde{p}_{ij} = \begin{cases} \left(\frac{x_{ji1}}{M}, \frac{x_{ji2}}{M}, \frac{x_{ji3}}{M} \right), & M = \max_j x_{ji3}, \text{ } CR_i \text{ criterio de beneficio} \\ \left(\frac{m}{x_{ji3}}, \frac{m}{x_{ji2}}, \frac{m}{x_{ji1}} \right), & m = \min_j x_{ji1}, \text{ } CR_i \text{ criterio de coste} \end{cases} \quad (4.6)$$

Mediante este proceso de normalización el rango de valores de los números fuzzy triangulares se restringen al intervalo [0, 1].

4.2.2.3 Proceso de decisión

Considerando la importancia de cada criterio, se puede construir la matriz de decisión fuzzy promediada y normalizada como sigue,

$$\tilde{V} = [\tilde{v}_{ji}]_{o \times n} \text{ donde } i=1,2,\dots,n \text{ y } j=1,2,\dots,o \text{ y } \tilde{v}_{ji} = \tilde{p}_{ji}(\cdot) \tilde{z}_i \quad (4.7)$$

Los elementos de la matriz anterior siguen siendo números fuzzy triangulares positivos y normalizados en el rango [0,1]. Las soluciones ideales positiva y negativa vienen dadas por la siguiente expresión,

$$A^+ = (\tilde{v}_1^+, \tilde{v}_2^+, \dots, \tilde{v}_n^+) \text{ y } A^- = (\tilde{v}_1^-, \tilde{v}_2^-, \dots, \tilde{v}_n^-) \quad (4.8)$$

donde $\tilde{v}_i^+ = (1,1,1)$ y $\tilde{v}_i^- = (0,0,0)$ con $i=1,2,\dots,n$.

La distancia de cada alternativa con respecto a A^+ y A^- se calcula como,

$$d_j^+ = \sum_{i=1}^n d(\tilde{v}_{ji}, \tilde{v}_i^+) \text{ y } d_j^- = \sum_{i=1}^n d(\tilde{v}_{ji}, \tilde{v}_i^-) \quad j=1,2,\dots,o \quad (4.9)$$

donde $d(\cdot, \cdot)$ es la distancia entre dos números fuzzy. El *método del vértice*, definido en Chen (2000), permite calcular la distancia entre dos números fuzzy según la ecuación

(4.10); dados $\tilde{a}=(a_1, a_2, a_3)$ y $\tilde{b}=(b_1, b_2, b_3)$ dos números triangulares fuzzy positivos.

Su distancia resulta ser,

$$d(\tilde{a}, \tilde{b}) = \sqrt{\frac{1}{3} \left[(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2 \right]} \quad (4.10)$$

El objetivo final del problema de decisión consiste en seleccionar una alternativa, con tal propósito, se define el coeficiente de proximidad a partir de las distancias definidas en la ecuación (4.9) obtenidas para cada alternativa A_j ($j = 1, 2, \dots, o$) con respecto a las respectivas distancias ideal positiva y negativa. Dicho coeficiente viene definido en Chen (2000) como,

$$cp_j = \frac{d_j^-}{d_j^+ + d_j^-} \quad \text{con } j=1, 2, \dots, o \quad (4.11)$$

Suponiendo que una alternativa real coincide exactamente con el valor ideal positivo, entonces d_j^+ es cero (distancia mínima) y d_j^- es la unidad (distancia máxima), en cuyo caso cp_j es igual a la unidad. En el otro extremo, cuando una alternativa toma el valor ideal negativo ocurre lo contrario, es decir d_j^- es cero y d_j^+ es la unidad, en cuyo caso cp_j toma el valor nulo. Esto implica que cuanto más se aproxime cp_j a uno, tanto más próximo está A_j de la solución ideal positiva y más alejado de la negativa. Por tanto, de acuerdo con el coeficiente de proximidad, se puede determinar el orden de preferencia de las alternativas y seleccionar la mejor de entre todas ellas según el criterio del máximo valor de dicho coeficiente. En resumen, el algoritmo de decisión para el caso multi-alternativa y multi-criterio para la toma de decisiones en el contexto fuzzy se concreta en los pasos siguientes:

- 1) Identificar los criterios CR_i para la elección de alternativas A_j .
- 2) Elegir las variables lingüísticas apropiadas.
- 3) Obtener las valoraciones fuzzy \tilde{x}_{ji} de cada alternativa A_j con respecto al criterio CR_i ; y determinar los pesos específicos o importancias relativas \tilde{z}_i concedidas a cada uno de los criterios.
- 4) Construir la matriz de decisión fuzzy \tilde{D} y la matriz fuzzy \tilde{P} normalizada.

- 5) Construir la matriz de decisión fuzzy promediada normalizada con los pesos asociados a cada criterio, \tilde{V}
- 6) Determinar las soluciones ideales positiva y negativa fuzzy, A^+ y A^- .
- 7) Calcular la distancia de cada alternativa a las soluciones ideales positiva y negativa, d_j^+ y d_j^- respectivamente.
- 8) Calcular el coeficiente de proximidad de cada alternativa, cp_j .
- 9) Ordenar las alternativas según el coeficiente de proximidad y seleccionar la mejor alternativa como aquella con el máximo valor de dicho coeficiente.

4.2.3 Cálculo de las competencias de los clasificadores individuales

Como se deduce de la teoría relativa al paradigma *MCDMF* en la sección previa, los pesos asignados a cada criterio son uno de los factores determinantes en la decisión final. En el caso que nos ocupa, los criterios con los que se van a seleccionar las alternativas son las salidas proporcionadas por los diferentes clasificadores en la fase de decisión, pues bien, en función de los valores de dichas salidas y de la importancia relativa asignada a cada clasificador así será la decisión final sobre la pertenencia de una muestra x a cada clase, donde ahora y bajo este paradigma las clases determinan las alternativas, como se verá posteriormente. El objetivo de esta sección consiste en determinar las *competencias* o *relevancias* relativas de cada clasificador individual que interviene en la combinación, a partir de las cuales se obtienen los pesos asignados a cada clasificador y de aquí los pesos específicos para cada criterio.

En el capítulo tres, se definen a partir de la ecuación (3.44) las competencias absolutas para cada clasificador atendiendo a la variación de los valores dados por las funciones criterio allí definidas; resultando ser finalmente las dos siguientes por las razones allí esgrimidas: la suma del error al cuadrado (*SE*) y la mínima varianza (*VC*). Dicha ecuación se reproduce aquí particularizada para cada una de las mencionadas funciones criterio,

$$\alpha^{SE} = 1 - \hat{\Delta}^{SE}(\hat{c}) \text{ y } \alpha^{VC} = 1 - \hat{\Delta}^{VC}(\hat{c}) \quad (4.12)$$

donde $\hat{\Delta}^{SE}(\hat{c})$ y $\hat{\Delta}^{VC}(\hat{c})$ se refieren a la variación relativa de sendas funciones *SE* y *VC*, que se obtienen a partir de las ecuaciones (3.42) y (3.43).

Como bien es sabido, durante la fase de entrenamiento, comenzamos con la observación de un conjunto X de n muestras de entrenamiento, esto es: $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \in \mathfrak{R}^3$. Cada muestra debe ser asignada a una clase determinada w_j , donde el número de clases posible es c , y por tanto $j = 1, 2, \dots, c$.

Pues bien, el conjunto de datos de entrenamiento X se divide en B subconjuntos $\{S_1, S_2, \dots, S_B\}$, tales que, $S_i \subset X$ y $S_i \cap S_k = \emptyset$, $\forall i \neq k$. Para cada subconjunto S_i , se obtiene $\hat{\Delta}^{SE}(c)$, según la ecuación (3.43) y además se determina cuál es la partición considerada como válida siguiendo el procedimiento descrito en la sección 4.2.1; G en la ecuación (3.43) se fija ocho, dado que este número es suficiente para todos los datos analizados en este trabajo de investigación. La partición se valida cuando se cumple que $\hat{\Delta}^{SE}(c) < U$, que es el criterio utilizado para estimar una partición aceptable. De este modo, el valor de c que cumple lo anterior se toma como el número de clases óptimo \hat{c}_i para el subconjunto S_i , siendo por tanto, en este caso $c = \hat{c}_i$. El umbral U se determina experimentalmente mediante el procedimiento descrito en el capítulo cinco, sección 5.3.1.5, anticipando ya que su valor resulta ser 0.1 según nuestros experimentos.

De este modo se obtiene un número de clases óptimo, según el criterio previo, para cada uno de los subconjuntos S_i . La mejor partición o equivalentemente el mejor número de clases para X se determina mediante votación mayoritaria a partir de los \hat{c}_i . El resultado de esta votación proporciona el número de clases óptimo \hat{c} para el conjunto completo X . Para cada subconjunto S_i con $\hat{c}_i \neq \hat{c}$ se determina una nueva partición, si bien esta vez asumiendo que el número de clases es exactamente \hat{c} . De esta forma, finalmente sobre el conjunto total X se establece una partición de acuerdo con el número estimado de clases \hat{c} . Con esta partición final se determinan los parámetros de aprendizaje derivados del proceso de entrenamiento para cada clasificador individual, según se describe en la sección 3.3, los cuales son almacenados en la BC , figuras 3.1 y 4.1, tras este proceso.

Por otra parte, a partir de los \hat{c}_i en cada subconjunto S_i se calculan las competencias para cada clasificador tal y como se describe a continuación. Dado que los clasificadores individuales FC , BP , SO , $L1$ y $L2$ estiman los centros de las clases según

sus respectivos procesos de entrenamiento, para cada uno de ellos se calcula $\alpha_l^{SE}(\hat{c}_i) = 1 - \hat{\Delta}^{SE}(\hat{c}_i)$ de acuerdo con la ecuación (4.12) y teniendo en cuenta los centros proporcionados por cada clasificador. El hecho de utilizar la función criterio SE para estos clasificadores es debido a que esta función obtiene sus valores precisamente a partir de los centros de las clases. El subíndice l se refiere a uno de los mencionados clasificadores, por tanto, $l = FC, BP, SO, LI$ y $L2$.

Por otro lado, el clasificador BN estima las funciones de densidad de probabilidad para cada clase teniendo en cuenta las diferencias relativas entre las muestras de la clase. Este hecho sugiere la utilización de la función criterio VC , que tiene en cuenta la dispersión de las muestras en las clases, con ello se obtiene $\alpha_{BN}^{VC}(\hat{c}_i) = 1 - \hat{\Delta}^{VC}(\hat{c}_i)$ también según la ecuación (4.12).

Los valores $\alpha_l^{SE}(\hat{c}_i)$ y $\alpha_{BN}^{VC}(\hat{c}_i)$ determinan los grados de competencia absoluta de cada clasificador individual y para cada subconjunto S_i . El máximo/mínimo grado de competencia absoluta se logra con los valores 1 y 0 respectivamente. Los valores finales de las *competencias absolutas* para cada clasificador se calculan a partir de la ecuación (4.13) teniendo en cuenta los subconjuntos S_i junto con su número de clases estimado \hat{c}_i ,

$$\alpha_m = \left| \sum_{i=1}^B (-1)^s \alpha_m^f(\hat{c}_i) \right| \quad (4.13)$$

donde m identifica cada clasificador como $FC, BP, SO, LI, L2$ o BN ; f es VC para BN y SE para los clasificadores restantes; con $s = 0$ si $\hat{c}_i = \hat{c}$ y $s = 1$ si $\hat{c}_i \neq \hat{c}$. En la expresión (4.13) se incluyen los términos de recompensa y penalización dependiendo de las condiciones expuestas previamente, de suerte que cuando se acierta con el número de clases se aplica la recompensa y cuando se produce un error se penaliza. Esto se consigue mediante el exponente s , que determina el signo positivo o negativo según se trate de un acierto o un error respectivamente.

4.2.4 Fase de decisión: combinación de clasificadores

Refiriéndonos de nuevo a la figura 4.1, ante la llegada de una nueva muestra o patrón \mathbf{x}_s , el problema que se plantea ahora consiste en tomar una decisión sobre la clasificación de dicha muestra en cada una de las clases disponibles y establecidas previamente durante la fase de entrenamiento.

Como bien es sabido, cada clasificador individual toma su propia decisión sobre la clasificación de dicha muestra según las salidas obtenidas en los correspondientes procesos de decisión descritos en el capítulo tres, sección 3.3, que sintetizamos de nuevo aquí por simplicidad. Así, *FC* proporciona los grados de pertenencia de dicha muestra a cada una de las clases; tanto *BP* como *BN* generan probabilidades de pertenencia de la muestra a cada una de las clases; *SO*, *L1* y *L2* proporcionan distancias de la muestra a los centros de las clases. Los tres primeros toman la decisión en función de los valores máximos de sus salidas, mientras que los tres últimos lo hacen en función de los valores mínimos.

Pues bien, la idea que subyace a la hora de aplicar el paradigma *MCDMF* consiste en tomar todos esos valores de salida (grados de pertenencia, probabilidades y distancias) con el fin de combinarlos siguiendo las directrices de dicho paradigma expuestas en la sección 4.2.2 y siguiendo los trabajos de Wang y Fenton (2006), Gu y Zhu (2006) y Chen (2000). La adaptación del proceso allí descrito se expone a continuación.

Partimos del hecho previo de que el número de clases ha sido estimado durante la fase de entrenamiento, resultando ser \hat{c} . Por simplicidad, de aquí en adelante dicho número queda renombrado como c , por tanto las clases disponibles donde la muestra \mathbf{x}_s debe ser clasificada son w_j con $j = 1, 2, \dots, c$.

Como se ha mencionado previamente, desde el punto de vista de la teoría de la decisión, el problema consiste en determinar a qué clase pertenece \mathbf{x}_s , esto supone elegir una de entre varias alternativas donde cada clase representa una de tales opciones. Más específicamente, desde el punto de vista de la teoría *MCDMF*, la elección de una clase es equivalente a elegir una alternativa. Por tanto, a partir de este momento se identifican las clases con las alternativas, esto es $w_j \equiv A_j$.

Por otro lado, la elección de una alternativa debe tomarse según ciertos *criterios* previamente establecidos. Como se ha mencionado previamente, los clasificadores individuales toman sus decisiones en función de las salidas obtenidas en la fase de decisión. Por este motivo, los criterios para elegir una alternativa bajo el paradigma *MCDMF* vienen determinados por las salidas proporcionadas por dichos clasificadores. Siguiendo la teoría expuesta en la sección 4.2.2 existen dos tipos de criterios, a saber: de *beneficio* y de *coste*. En los primeros, los valores mayores representan la mejor opción,

mientras que en los segundos la mejor opción viene dada por los valores más bajos. Pues bien, según lo expuesto anteriormente los clasificadores FC , BP y BN toman sus decisiones en base a los valores máximos de sus salidas y SO , $L1$ y $L2$ lo hacen en función de los mínimos. Esta reflexión nos sugiere considerar los dos grupos de clasificadores por separado, a partir de los cuales se derivarán dos criterios, como se explica seguidamente. Teniendo en cuenta los números triangulares fuzzy, definimos los dos criterios siguientes: el criterio CR_1 de *beneficio* y el criterio CR_2 de *coste*. Ambos representan tres números triangulares fuzzy de forma que en el caso de CR_1 sus tres componentes son los valores de salida proporcionados por FC , BP y BN y en el caso de CR_2 sus componentes son los valores de salida dados por SO , $L1$ y $L2$.

Siguiendo los pasos del algoritmo descrito al final de la sección 4.2.2, a continuación se formaliza la adaptación de nuestro problema al marco del paradigma *MCDMF*. En lo que sigue nos referiremos a dicho algoritmo de forma genérica siguiendo exactamente esos mismos pasos.

- 1) El primer paso del algoritmo consiste en la *identificación de los criterios*. Pues bien los criterios seleccionados son CR_1 y CR_2 según lo expresado previamente.
- 2) El segundo paso consiste en la elección de las *variables lingüísticas*. Atendiendo a los criterios seleccionados, éstas se definen como sigue:
 - 2.1) CR_1 toma los valores según el vector $X_s^j \equiv \{\mu_{sj}, P_b(w_j / x_s), P_p(w_j / x_s)\}$, cuyas componentes son las salidas proporcionadas por FC , BP y BN respectivamente, tabla 3.1.
 - 2.2) CR_2 toma los valores según el vector $Y_s^j \equiv \{d(x_s, v_j^0), d(x_s, v_j^1), d(x_s, v_j^2)\}$; donde $d(x_s, v_j^r)$ representa la distancia Euclídea entre la muestra x_s y el centro de cada clase w_j representados por v_j^0 , v_j^1 y v_j^2 para SO , $L1$ y $L2$ respectivamente, según la tabla 3.1. Obsérvese, que se ha cambiado la nomenclatura para designar estos centros con respecto a la utilizada para los mismos centros en el capítulo tres, sección 3.3 y la misma tabla 3.1.
- 3) El tercer paso consta de dos partes, la primera consiste en la obtención de las valoraciones fuzzy de cada alternativa con respecto a cada criterio. Esto se lleva

a cabo mediante la definición de los siguientes números triangulares fuzzy:

$$X_s^j \equiv \{X_{s1}^j, X_{s2}^j, X_{s3}^j\} \text{ con } X_{s1}^j \leq X_{s2}^j \leq X_{s3}^j \text{ e } Y_s^j \equiv \{Y_{s1}^j, Y_{s2}^j, Y_{s3}^j\} \text{ con } Y_{s1}^j \geq Y_{s2}^j \geq Y_{s3}^j.$$

La segunda parte del tercer paso implica la obtención de la importancia relativa de cada criterio, que desemboca finalmente en el cómputo de los pesos específicos asociados. Esta tarea se lleva a cabo a partir de las competencias absolutas α_m definidas en la ecuación (4.13), cuyo cómputo se obtiene promediando para cada grupo de clasificadores según los criterios CR_1 y CR_2 . En efecto, para CR_1 la competencia absoluta promediada α^1 resulta ser el valor medio de α_{FC} , α_{BP} y α_{BN} ; mientras que para CR_2 la competencia absoluta promediada α^2 es el valor medio entre α_{SO} , α_{L1} y α_{L2} . Finalmente, los pesos específicos asignados a cada criterio se calculan como $\tilde{z}_1 = \alpha^1 / (\alpha^1 + \alpha^2)$ y $\tilde{z}_2 = \alpha^2 / (\alpha^1 + \alpha^2)$ para CR_1 y CR_2 respectivamente, cumpliendo la restricción de que $\tilde{z}_1 + \tilde{z}_2 = 1$; por comodidad renombramos estos pesos como z_1 y z_2 .

- 4) El cuarto paso consiste en la construcción de la matriz de decisión fuzzy y la matriz fuzzy normalizada. Esta tarea se lleva a cabo a partir de la ecuación (4.6). Como se ha mencionado previamente, en nuestro enfoque las alternativas se materializan en la elección de las clases w_j ; de forma que para llevar a cabo la normalización es preciso obtener los siguientes valores: $M = \max_j \{X_{s3}^j\}$ y $m = \min_j \{Y_{s3}^j\}$.
- 5) El quinto paso del algoritmo se concreta a partir de las matrices promediadas y normalizadas obtenidas en los pasos previos, cuyos elementos están dados por X_{sM}^j y Y_{sN}^j . En la tabla 4.1 se muestra la matriz de decisión normalizada y promediada.
- 6) El sexto paso se refiere a la definición de las soluciones ideales positivas y negativas. En este caso el propio proceso de normalización, tanto para el criterio de beneficio como de coste, restringe directamente los valores al rango $[0,1]$, siendo las mejores elecciones las que se toman con valores próximos a 1. Por este motivo, las soluciones ideales positivas y negativas son exactamente las definidas a través de la ecuación (4.8), es decir $\tilde{v}_i^+ = (1,1,1)$ y $\tilde{v}_i^- = (0,0,0)$.

Tabla 4.1 Tabla de decisión normalizada y promediada según los criterios y sus pesos asociados

		<i>criterios (pesos)</i>	
		$CR_1 (z_1)$	$CR_2 (z_2)$
<i>decisión</i>	w_1	$X_{sM}^1 = [X_{s1}^1/M, X_{s2}^1/M, X_{s3}^1/M] z_1$	$Y_{sN}^1 = [m/Y_{s1}^1, m/Y_{s2}^1, m/Y_{s3}^1] z_2$
	w_2	$X_{sM}^2 = [X_{s1}^2/M, X_{s2}^2/M, X_{s3}^2/M] z_1$	$Y_{sN}^2 = [m/Y_{s1}^2, m/Y_{s2}^2, m/Y_{s3}^2] z_2$

	w_c	$X_{sM}^c = [X_{s1}^c/M, X_{s2}^c/M, X_{s3}^c/M] z_1$	$Y_{sN}^c = [m/Y_{s1}^c, m/Y_{s2}^c, m/Y_{s3}^c] z_2$

- 7) El séptimo paso consiste en el cálculo de las distancias de cada alternativa a las soluciones ideales positiva y negativa. Esto se realiza por medio de la ecuación (4.9).
- 8) El octavo paso se realiza calculando el coeficiente de proximidad para cada alternativa a través de la ecuación (4.11).
- 9) Finalmente, el noveno y último paso consiste en elegir la mejor alternativa, es decir la clase w_j más apropiada para la muestra x_s . Dado que en nuestro modelo, alternativas y clases son equivalentes, la mejor clase w_j para x_s se corresponde con la alternativa que proporciona el máximo valor del coeficiente de proximidad, es decir la elección se hace de acuerdo con la regla siguiente,

$$w_j = \arg \max_j \{ cp_j \} \quad (4.14)$$

4.3 Diseño de un clasificador local no supervisado: Integral Fuzzy

En el esquema del clasificador local no supervisado basado en la *Integral Fuzzy (IF)* se utilizan los mismos seis clasificadores individuales que en el caso de la estrategia *MCDMF*. Además, la fase de entrenamiento es idéntica a la descrita en la sección 4.2.1. Por tanto, nos centraremos en la descripción de la fase de decisión, que es donde difieren sustancialmente *IF* y *MCDMF*. El proceso de decisión se lleva a cabo mediante el algoritmo basado en la integral fuzzy descrito en el capítulo 3, sección 3.5.2.5.

Dada la muestra \mathbf{x}_s a clasificar y con el fin de restringir los valores de salida de los seis clasificadores individuales al rango $[0,1]$, se lleva a cabo el mismo proceso de normalización que el descrito en la sección 4.2.4, tabla 4.1. A partir de los valores normalizados se construye la tabla 4.2, que es una fiel reproducción de la 4.1 a la que se ha añadido la columna final, conteniendo las densidades fuzzy para cada clasificador, requeridas por el algoritmo de la integral fuzzy.

Tabla 4.2 Salidas de los clasificadores y densidades fuzzy asociadas

<i>Clasificador</i>	<i>Clases</i>				<i>densidades fuzzy</i>
	w_1	w_2	\dots	w_c	
<i>FC</i>	X_{s1}^1/M	X_{s1}^2/M	\dots	X_{s1}^c/M	g^{FC}
<i>BP</i>	X_{s2}^1/M	X_{s2}^2/M	\dots	X_{s2}^c/M	g^{BP}
<i>BN</i>	X_{s3}^1/M	X_{s3}^2/M	\dots	X_{s3}^c/M	g^{BN}
<i>SO</i>	m/Y_{s1}^1	m/Y_{s1}^2	\dots	m/Y_{s1}^c	g^{SO}
<i>L1</i>	m/Y_{s2}^1	m/Y_{s2}^2	\dots	m/Y_{s2}^c	g^{L1}
<i>L2</i>	m/Y_{s3}^1	m/Y_{s3}^2	\dots	m/Y_{s3}^c	g^{L2}

Para poder aplicar el algoritmo de la integral fuzzy es necesario identificar la procedencia de los valores X e Y normalizados en relación al clasificador que los ha originado. Sin pérdida de generalidad supóngase la asignación dada en la tabla 4.2. Es decir, X_{s1} , X_{s2} y X_{s3} con *FC*, *BP* y *BN* respectivamente e Y_{s1} , Y_{s2} y Y_{s3} con *SO*, *L1* y *L2*. Además, para cada clasificador es necesario obtener las densidades fuzzy, por lo cual recurrimos de nuevo a las competencias obtenidas para cada clasificador según la ecuación (4.13), de forma que dado un m cualquiera de los seis clasificadores, $g^m = \alpha_m / \sum_{k \in \Omega} \alpha_k$ con $\Omega = \{FC, BP, BN, SO, L1, L2\}$. A partir de este momento se aplica exactamente el algoritmo de la integral fuzzy descrito en el capítulo tres, sección 3.5.2.6. Es necesario establecer las siguientes correspondencias entre las tablas 3.2 y 4.2; así, los L clasificadores de la tabla 3.2 son ahora los seis clasificadores individuales de la tabla 4.2, de forma que $i \equiv \Omega = \{FC, BP, BN, SO, L1, L2\}$ y por tanto los coeficientes de la matriz perfil de decisión $d_{ij}(\mathbf{x}_s)$, definidos en la ecuación (3.49), que aparecen en la tabla 3.2 son exactamente los definidos en las celdas equivalentes de la tabla 4.2, por ejemplo $d_{1,1}(\mathbf{x}_s) \equiv X_{s1}^1/M$ o $d_{L,c}(\mathbf{x}_s) \equiv m/Y_{s3}^c$ por referir dos de ellos.

4.4 Diseño de un clasificador global no supervisado: Mapas Cognitivos Fuzzy

En la figura 4.2 se muestra el diseño del clasificador no supervisado basado en el paradigma de los Mapas Cognitivos Fuzzy, *MCF*. Como todo clasificador, el procedimiento posee las ya conocidas dos fases de entrenamiento y decisión. La fase de entrenamiento está diseñada de una forma totalmente similar a la misma fase del clasificador basado en el paradigma *MCDMF*; como allí, es en la fase de decisión donde se lleva a cabo el proceso de combinación de los clasificadores individuales.

En primer lugar, se describe brevemente, el mecanismo de funcionamiento del clasificador durante la fase de entrenamiento, sección 4.4.1, para posteriormente incidir en la fase de decisión, sección 4.4.3. Previamente, se establecen los conceptos teóricos básicos relativos a la teoría de los *MCFs*, sección 4.4.2, que constituyen el fundamento de la combinación, tras lo cual se formula la propuesta de combinación en el contexto de la clasificación de texturas en imágenes naturales. Para ello, se tienen en cuenta los principios de la Gestalt, expuestos en la introducción, en el caso concreto que nos ocupa se seleccionan: *proximidad, similitud, continuidad y conectividad*.

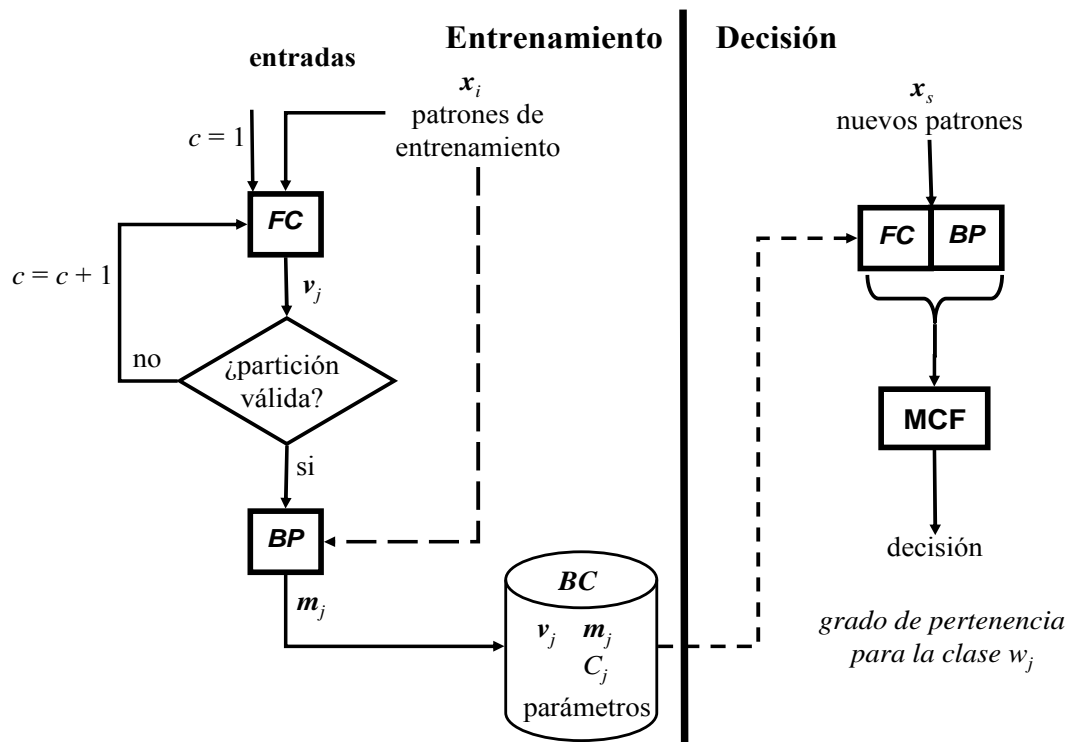


Figura 4.2 Arquitectura de un clasificador híbrido basado en el enfoque MCF: entrenamiento y decisión

4.4.1 Fase de entrenamiento

La fase de entrenamiento sigue un procedimiento similar al descrito en la sección 4.2.1 para el MCDMF, en lo que respecta al establecimiento de las clases iniciales y validación de la partición. Se trata pues del mismo diseño, de forma que como en aquél, el proceso llega a ser no supervisado, aún partiendo del clasificador no supervisado *FC*. Tras la validación de la partición, ésta se pasa al estimador de máxima verosimilitud paramétrico (*BP*), que inicia su correspondiente proceso de aprendizaje. El proceso de validación es igual que el llevado a cabo para el MCDMF, la diferencia fundamental estriba en el hecho de que aquí se entrenan dos clasificadores frente a los seis entrenados en el MCDMF.

Tras los respectivos procesos de entrenamiento, tanto *FC* como *BP* almacenan en la *BC* sus correspondientes parámetros aprendidos, quedando éstos disponibles para su posterior recuperación durante la fase de decisión con el fin de proceder a la combinación de los clasificadores.

4.4.2 Mapas Cognitivos Fuzzy: descripción general

Antes de abordar la fase de decisión, es necesario proceder a la descripción general del paradigma de los *MCFs*. La elección de este paradigma se justifica por su capacidad para incorporar la información espacial existente en las imágenes y también, de forma general, porque proporciona un buen mecanismo de decisión incluso en procesos complejos (Bueno y Salmerón, 2009; Koulourioti y col., 2005). En su concepción original, fueron los mapas cognitivos (Axelrod, 1976), planteándose posteriormente bajo la perspectiva fuzzy (Kosko, 1992); de este modo, los *MCFs* se estructuran como grafos dirigidos con capacidad de retroalimentación. Cada nodo del grafo se conoce como *concepto causal*. Dos conceptos causales C_i y C_j se encuentran conectados mediante un arco dirigido o conexión causal. Esta conexión tiene asociado un peso causal s_{ij} , que mide cómo el concepto C_i afecta (causa) al concepto C_j . Un valor de $s_{ij} = 0$ expresa ausencia de causalidad, mientras que valores cumpliendo $s_{ij} > 0$, indican causalidad creciente, lo que se traduce en que C_j aumenta en la medida que también aumenta C_i . Por contra, cuando $s_{ij} < 0$, lo que se expresa es una causalidad decreciente o negativa, es decir C_j disminuye cuando C_i aumenta, y viceversa, C_j aumenta cuando C_i

disminuye. Un típico ejemplo lo propone Kosko (1992) cuando sostiene que la congestión del tráfico rodado en cualquier vía junto con una climatología adversa aumentan la siniestralidad en dicha vía.

Los *MCFs* más simples poseen valores de conexión en el conjunto discreto $\{-1, 0, 1\}$, en cuyo caso, si ocurre una causalidad tomará el máximo valor positivo o negativo. En la figura 4.3 se muestra un sencillo esquema de nodos interconectados.

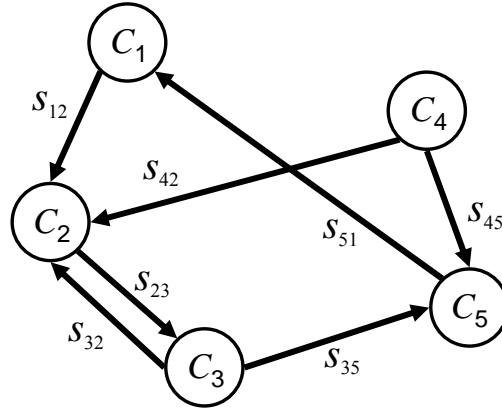


Figura 4.3 Un ejemplo sencillo de Mapa Cognitivo fuzzy

Según el ejemplo de la figura 4.3, se puede definir la matriz de adyacencia de dimensión 5×5 , a partir de las relaciones de causalidad entre los nodos según la ecuación (4.15), donde se observa que el nodo C_2 ejerce sobre el C_3 una influencia negativa, mientras que el resto de los nodos con conexiones ejercen influencias positivas. En los demás casos, no existen conexiones entre los nodos.

$$E = \begin{matrix} & \begin{matrix} C_1 & C_2 & C_3 & C_4 & C_5 \end{matrix} \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix} \quad (4.15)$$

Los conceptos pueden representar procesos, eventos, valores o estados, entre otras categorías. En nuestro modelo combinado serán estados de los nodos en una red neuronal.

La generalización de las ideas previas resulta relativamente sencilla bajo el paradigma de los *MCFs*. En efecto, los conceptos pueden considerarse estados tomando valores en

el intervalo $[-1,+1]$; por otro lado, las conexiones indican el grado de influencia ejercido por el concepto causal C_i sobre el C_j tomando igualmente valores en el intervalo $[-1,+1]$ (Tsardias and Margaritis, 1997, 1999; Kosko, 1986,1992; Miao and Liu, 2000).

El modelado de sistemas dinámicos utilizando los *MCFs* tiene varias ventajas. La más importante es que son estructuras relativamente simples e intuitivas. Se caracterizan también por su fácil adaptabilidad a un dominio determinado con alta capacidad de englobar los factores de decisión (Bueno y Salmerón, 2009). En suma, se puede decir que permiten realizar abstracciones con una capacidad de razonamiento alta (Stach y col., 2005). Además, otra ventaja de modelar sistemas dinámicos utilizando un *MCF* proviene del hecho de que aunque el estado inicial de los conceptos sea incompleto o incorrecto, el mecanismo de actualización puede completar o corregir estas deficiencias iniciales (Sharif e Irani, 2005).

Dado un *MCF* con n conceptos C_i ($i = 1, \dots, n$) y asignando los conceptos a nodos, bajo la perspectiva de las redes neuronales, el valor de cada concepto en una iteración dada $t+1$, puede obtenerse a partir de los valores de los conceptos en la iteración anterior según se expresa en la ecuación (4.16) (Martchenko y col., 2003),

$$A_i(t+1) = f \left(s_{ii}A_i(t), \sum_{\substack{k=1 \\ k \neq i}}^n s_{ki}A_k(t) \right) = f(x, y) \quad (4.16)$$

donde:

1. $A_i(t+1)$ es el valor del concepto C_i del nodo asociado en la iteración $t + 1$,
2. $A_i(t)$ es el valor del concepto C_i del nodo asociado en la iteración t ,
3. s_{ki} representa el peso causal fuzzy de la conexión entre los dos conceptos dados C_k y C_i , según se ha definido previamente,
4. s_{ii} representa el peso causal fuzzy de auto-conexión para el concepto C_i ,
5. $f(x,y)$ es una función de umbral, que transforma el resultado de la operación a valores de salida en un intervalo fuzzy dado, donde se expresan los valores de los

conceptos. La transformación definida en la ecuación (4.16) realmente lo que realiza es una actualización del valor del concepto de cada nodo en función de los valores de los conceptos previos y de las conexiones causales entre conceptos a través de sus pesos. La función f puede definirse de diferentes maneras con tal de que proporcione los valores de salida dentro del intervalo fuzzy especificado. Existen en la literatura diversas definiciones para f , según la aplicación; entre ellas destacan la tangente hiperbólica (Marchenko y col. 2003; Yu y Tzeng, 2006) o la que se utiliza en este trabajo, definida en la ecuación (4.18) y elegida por las razones allí expuestas.

Volviendo sobre la ecuación (4.16) e identificando términos con respecto a la función $f(x,y)$, la variable x expresa la influencia ejercida por el concepto i sobre sí mismo durante la actualización de su valor, mientras que y determina la influencia ejercida por los k conceptos conectados con i .

En la siguiente sección, correspondiente a la fase de decisión del clasificador, se describe la adaptación del paradigma *MCF* para la combinación de los dos clasificadores base *FC* y *BP* utilizados en el esquema descrito en la figura 4.2, bajo el contexto de la clasificación de las texturas naturales en imágenes de exterior.

4.4.3 Fase de decisión

Por similitud entre la topología de los *MCFs* y las redes neuronales, de aquí en adelante nos referiremos a los conceptos causales como nodos, de esta forma el concepto C_i se identifica con el nodo i . Por otra parte y bajo esta misma perspectiva, en lugar de referirnos a valores de los conceptos, diremos valores o estados de los nodos indistintamente, para expresar la misma idea.

Para la adaptación de los *MCFs* al contexto de la clasificación de imágenes tenemos en cuenta que cada píxel de la imagen a clasificar, se identifica como un nodo. Por consiguiente, cada nodo será engendrado por un píxel de dicha imagen; de este modo, un píxel en la posición espacial (x,y) generará un nodo i , estableciéndose una asociación biunívoca entre nodo y píxel de la forma $i \equiv (x, y)$.

Dada una nueva muestra (píxel) a clasificar representada por su vector de propiedades \mathbf{x}_s , el problema consiste en decidir a qué clase w_j pertenece de entre las identificadas

durante la fase de entrenamiento. Las tres componentes del vector \mathbf{x}_s se corresponden con los tres valores espectrales de las componentes de color, en el modelo RGB, del píxel o nodo que representa $i \equiv (x, y)$. Una importante observación, que conviene hacer aquí, es la relacionada con las localizaciones espaciales entre un píxel y sus vecinos, ya que el modelo basado en los *MCFs* tiene su fundamento en las relaciones espaciales de vecindad, que permitirán la aplicación de los principios de la Gestalt mencionados en la introducción.

4.4.3.1 Topología de los *MCFs*

Dadas las w_j clases posibles construiremos una red de nodos por cada clase que designamos como red_j , siendo por tanto $j = 1, \dots, c$. Esto significa que tendremos tantas redes como clases disponibles. Cada nodo i en la red_j se asocia con el píxel localizado en la posición (x, y) de la imagen a clasificar, según se ha establecido previamente. Por consiguiente, siendo el nodo (píxel) $i \equiv (x, y)$ el objetivo de la clasificación, la decisión tomada por el clasificador propuesto estará basada en los valores de los *estados* de los nodos en las diferentes redes del tipo red_j , como se verá posteriormente. Establecida la asociación biunívoca entre nodo y píxel, ambos términos se consideran equivalentes. De este modo, aunque siempre es el píxel el elemento a clasificar en la imagen, a veces nos referiremos a la clasificación del nodo sin ninguna distinción formal.

El estado del nodo i en cada red_j se inicializa asignando a dicho nodo el grado de pertenencia, μ_i^j , proporcionado por el clasificador individual *FC*, como resultado de la salida de su proceso de decisión individual según la ecuación (3.5) y tabla 3.1. Por conveniencia de notación, con respecto a dicha ecuación, hacemos $\mu_{sj} \equiv \mu_i^j$ donde el subíndice s hace referencia a la muestra \mathbf{x}_s , identificada aquí como el vector representante del nodo $i \equiv (x, y)$, j identifica la *red*, esto es, la clase w_j .

Por tanto, según lo anterior, dichos grados de pertenencia son los valores de los estados iniciales de los nodos en cada una de las redes; es decir, μ_i^j es el valor del estado del nodo i en la red_j . Volviendo de nuevo sobre la ecuación (4.16), los valores iniciales de los conceptos A_i son ahora los grados de pertenencia mencionados y asociados a los nodos, diferenciándose entre ellos según su ubicación en cada una de las redes mediante el superíndice j .

4.4.3.2 Criterio de decisión

Si nos detuviéramos en el estado inicial de las c redes donde los estados de los nodos son los grados de pertenencia a las respectivas clases, según el clasificador individual FC , podríamos utilizar la regla de decisión asociada a este clasificador, ecuación (3.7), para clasificar los píxeles de una imagen dada. No obstante, estos estados iniciales van a ser modificados mediante el proceso de actualización propio de los $MCFs$, como se verá posteriormente; de suerte que a la finalización del mismo, los nodos alcanzarán nuevos valores en sus estados, como consecuencia de dicho proceso. Los estados resultantes constituyen la base sobre la que se toman las decisiones finales, siendo válida la misma regla de decisión definida en (3.7), la cual formulada en el contexto de los $MCFs$, se expresa como: $i \in w_j$ si $\mu_i^j > \mu_i^h, \forall j \neq h$ con $j, h = 1, \dots, c$. Resumiendo, la decisión se toma según el máximo valor de los estados para el nodo i en las c redes.

4.4.3.3 Proceso de actualización

Los grados de pertenencia tal y como son proporcionados por el clasificador FC toman por definición valores restringidos en el rango $[0, +1]$, ecuación (3.5). Si bien, para su utilización bajo el esquema de los $MCFs$ propuesto, se escalan linealmente al rango $[-1, +1]$, con ello los estados iniciales de los nodos en las c redes toman valores en ese intervalo.

Durante el proceso de actualización, los valores de los estados de los nodos en cada red son reforzados o penalizados de forma iterativa según las influencias ejercidas por los otros nodos que mantienen conexiones con él. El objetivo consiste en facilitar la toma de decisiones basándonos en valores más estables de los estados, tratando de conseguir así una mejor clasificación de los píxeles en las imágenes. Es decir, se pretende obtener un método robusto cuya base es la estabilidad de las redes.

Volviendo sobre la ecuación (4.16) se observa la necesidad de definir los pesos causales, que en nuestro diseño deben distinguirse según la red_j , por el hecho de disponer de c redes; esos pesos se identifican como s_{ik}^j con valores en el intervalo causal fuzzy $[-1, +1]$, que materializan la influencia ejercida por el nodo k sobre el nodo i en la red_j , aumentando o disminuyendo el valor del estado de dicho nodo i según el mecanismo descrito en la sección 4.4.2.

Centrándonos en nuestro problema de clasificación de las texturas en imágenes, dada una imagen de tamaño $M \times N$, el objetivo consiste en clasificar cada píxel i localizado en (x,y) como perteneciente a una clase w_j . El número de nodos en la red_j coincide con el tamaño de la imagen, $n = M \times N$, distribuyéndose espacialmente de la misma manera que los píxeles lo hacen en la imagen.

Haciendo balance de la situación hasta el momento, se dispone de lo siguiente:

1. Los valores de los estados iniciales de los nodos i en cada red_j , se corresponden con los grados de pertenencia proporcionados en su salida por el clasificador FC durante la fase de decisión, esto es $\mu_i^j(t=0) = \mu_i^j$, donde el primer término se refiere a los valores de los estados en el instante inicial, esto es en $t = 0$ y el segundo expresa los grados de pertenencia proporcionados directamente por FC .
2. Dados los nodos i y k en la red_j , si existe una relación causal entre ambos, ésta posee asociado un peso causal de conexión dado por s_{ik}^j .
3. A partir del estado inicial se desencadena el proceso de actualización de los estados de los nodos en todas las redes siguiendo el esquema dado por la ecuación (4.16), de suerte que en cada iteración se modifican los estados de acuerdo con los mecanismos de recompensa y castigo establecidos, que abordaremos seguidamente.
4. Por tanto, en este momento, quedan por especificar dos cosas. Por un lado la función de umbral f , que aparece en la ecuación (4.16) y por otro determinar los pesos de conexión entre los nodos. El primer aspecto lo abordamos inmediatamente, mientras que el segundo se pospone para la sección 4.4.3.4. No obstante, y como consecuencia de la redefinición del proceso de actualización original definido en la ecuación (4.16) se introduce un nuevo término, que tiene que ver con un nuevo aspecto conceptual conocido como factor de decadencia, que se trata en la sección 4.4.3.5.

De acuerdo con los trabajos de Tsardias y Margaritis (1997,1999), el valor del estado del nodo i en la iteración $t + 1$ se obtiene según la ecuación (4.17).

$$\mu_i^j(t+1) = f(\mu_i^j(t), D_i^j) - d_i^j \mu_i^j(t) = f\left(\mu_i^j(t), \sum_{k=1}^n s_{ik}^j(t) \mu_k^j(t)\right) - d_i^j \mu_i^j(t) \quad (4.17)$$

Comparando la ecuación (4.17) con la (4.16) se pueden establecer las siguientes analogías y diferencias:

1. Se identifican las variables de la función f como $x = \mu_i^j(t)$ e $y = \sum_{k=1}^n s_{ik}^j(t) \mu_k^j(t)$ con la misma interpretación que en (4.16) salvo en lo que respecta a la distinción con respecto a cada red_j , razón por la cual en (4.17) se introduce el superíndice j en todos los términos. A_i es equivalente a $x = \mu_i^j$ con la distinción de j .
2. La diferencia fundamental surge por la introducción del *factor de decadencia*, expresado como $d_i^j \in [0,1]$ en el segundo término de (4.17). Este factor ha sido propuesto por Tsardias y Margaritis (1997,1999) y refleja la tendencia natural de los nodos a permanecer en un estado inactivo, similar a lo que podría ocurrir en los fenómenos biológicos, según expresan los citados autores. Por tanto, desde la perspectiva que nos ocupa, dicho factor determina la fracción a sustraer del valor del estado resultante como consecuencia de dicha intencionalidad. Cuanto mayor sea dicho factor más fuerte será el mecanismo de decadencia.

Siguiendo la propuesta formulada por Tsardias y Margaritis (1997, 1999) una función de umbral típica f , con resultados satisfactorios es precisamente la definida en la ecuación (4.18), que ha sido exhaustivamente estudiada en esos trabajos. Dicha función proviene del famoso sistema experto MYCIN para la agregación de los factores de certeza, cuya descripción original puede encontrarse en Shorliffe (1976) o Buchanan y Shorliffe (1984) entre otros.

$$f(x, y) = \begin{cases} x + y(1-x) & \text{si } x, y \geq 0, \\ x + y(1+x) & \text{si } x, y < 0, \\ (x+y)/(1-\min(|x|, |y|)) & \text{de otro modo} \end{cases} \quad (4.18)$$

donde los valores x e y deben satisfacer la restricción $|x|, |y| \leq 1$. Inicialmente, cuando t es cero, $x = \mu_i^j$ cumpliéndose dicha restricción por la propia definición de los grados de pertenencia y por su posterior linealización al rango $[-1, +1]$, según se indica

previamente. Sin embargo, para $y = D_i^j = \sum_{k=1}^n s_{ik}^j \mu_k^j(t)$ puede no cumplirse, ya que la influencia de los nodos k podría llevar a que dicha función tome valores fuera del rango $[-1, +1]$. Para que D_i^j tome valores en el rango requerido, Tsardiras y Margaritis (1999) proponen aplicarle la función sigmodal definida por la tangente hiperbólica, $D_i^j = \tanh(\sigma D_i^j)$, donde σ es una constante de suavizado, que permite controlar la influencia de los nodos vecinos sobre el nodo i , se ha elegido un valor de 0.8 para disminuir hasta ese nivel el grado de influencia total de estos nodos. Una vez hecho esto, la restricción previa exigida para x e y se cumple teniendo en cuenta sólo la función definida en (4.18). No obstante, puede ocurrir que debido al término de decadencia, los valores finales actualizados según la ecuación (4.17) incumplan tal restricción. Esto es, podría suceder que $\mu_i^j(t+1)$ sea mayor que +1 o menor que -1, lo que supondría que en la siguiente iteración, cuando se esté calculando $\mu_i^j(t+2)$ ocurra que $|x| > 1$. Para evitar esto, cuando se produzca tal situación, los valores finales actualizados de $\mu_i^j(t+1)$ se limitan al rango $[-1, +1]$, asignándosele el valor +1 ó -1 según que el límite se haya sobrepasado en sentido positivo o negativo respectivamente.

En resumen, el mecanismo iterativo definido en la ecuación (4.17) representa el proceso de actualización de los estados de los nodos. Volviendo sobre dicha ecuación, podemos realizar las dos observaciones siguientes de interés sobre la actualización del estado del nodo i en t , $\mu_i^j(t)$ para llegar al estado $\mu_i^j(t+1)$:

1. En parte dicha actualización se produce por la influencia de los nodos vecinos a través del término D_{ik}^j .
2. Por otra parte, la propia influencia del estado actual $\mu_i^j(t)$ también interviene en la actualización bajo dos formas, en la propia función f y conjuntamente con el factor de decadencia.

4.4.3.4 Definición de los pesos de conexión

Teniendo en cuenta la ecuación (4.17), en este momento abordamos el procedimiento para determinar los pesos causales fuzzy, $s_{ik}^j(t)$, que aparecen en dicha ecuación. Cada

uno de estos pesos se obtiene mediante la combinación de dos coeficientes que tienen en cuenta las influencias de los k nodos vecinos sobre el nodo i en la red_j , a saber:

a) coeficiente de *regularización*

b) coeficiente *contextual*

a) *Definición del coeficiente de regularización*

Se trata de un coeficiente dinámico, que varía con el número de iteraciones t , cuya definición viene dada por la siguiente expresión,

$$r_{ik}^j(t) = \begin{cases} 1 - |\mu_i^j(t) - p_k^j| & k \in N_i^m, i \neq k \\ 0 & k \notin N_i^m, i = k \end{cases} \quad (4.19)$$

Seguidamente explicamos cada uno de los términos involucrados en esta ecuación.

Como se ha mencionado previamente, cada red_j posee tantos nodos como píxeles posea la imagen a clasificar, o dicho de otro modo, cada píxel genera un nodo en la red_j . Dado que las imágenes mantienen el concepto espacial sobre la distribución de los píxeles, se define el concepto de vecindad de un nodo i en la red como el conjunto de nodos k que rodean espacialmente a i . Así pues, N_i^m representa los m nodos vecinos del nodo i . Desde el punto de vista de la imagen N_i^m es la región m -conectada, siendo una práctica habitual utilizar una región 8-conectada, considerando así un píxel y sus ocho vecinos; por tanto, en este trabajo se toma m igual a ocho. Se pueden definir otros tamaños de vecindad, si bien en los experimentos llevados a cabo en nuestras investigaciones, este valor ha resultado ser suficiente a juzgar por los resultados obtenidos, ya que valores mayores hacen que los píxeles vecinos ejerzan una influencia excesiva sobre el píxel i .

A partir de la ecuación (4.19), se observa que solamente los nodos m -conectados contribuyen al cómputo del coeficiente de regularización, mientras que la contribución de los nodos fuera de dicho entorno, se declara nula. Esto es una consecuencia inmediata de la aplicación de los principios de proximidad, similitud, continuidad y conectividad en el contexto de las imágenes, según la filosofía de la Gestalt e introducidos en la sección 4.1.

Continuando con la ecuación (4.19), se observa que en el cómputo del coeficiente de regularización para una iteración dada t , intervienen tanto $\mu_i^j(t)$ como p_k^j . El primero

define el estado del nodo i , que está bajo el proceso de actualización en la red_j . Por otra parte, p_k^j representa la probabilidad de que el nodo (píxel) k representado por el vector \mathbf{x}_k , cuyas componentes son sus valores espectrales en el modelo de color RGB, pertenezca a la clase w_j . Esta probabilidad, corresponde a la salida proporcionada por el clasificador individual BP como consecuencia de su propio proceso de decisión según la ecuación (3.9), es decir, en el caso que nos ocupa, $p_k^j \equiv P_b(w_j / \mathbf{x}_k)$, donde las razones sobre la inclusión del subíndice b se explican en la sección 3.3.2.2. Debido a que las probabilidades están definidas en el rango $[0, +1]$, para el cómputo del coeficiente de regularización se escalan dichos valores linealmente al intervalo $[-1, +1]$, coincidiendo así con el rango de valores posibles para los estados del nodo i , $\mu_i^j(t)$.

Con lo expuesto hasta aquí, se dice que existe un alto grado de compatibilidad entre el nodo i y su vecino k en la red_j cuando $\mu_i^j(t)$ y p_k^j son similares, lo que intuitivamente se corresponde con el hecho de que dichos nodos tienen también posibilidades similares de pertenecer a la misma clase según la información proporcionada por la red_j a través del valor del estado $\mu_i^j(t)$ y la información proporcionada por el clasificador BP mediante p_k^j . El coeficiente de regularización $r_{ik}^j(t)$ toma el valor máximo de $+1$, dentro de su rango de definición, cuando $\mu_i^j(t)$ y p_k^j son iguales y el mínimo valor de -1 cuando son totalmente opuestos, por ejemplo con $\mu_i^j(t) = +1$ y $p_k^j = -1$ simultáneamente. Cuanto mayor sea la proximidad de los valores a -1 menor resulta la posibilidad de pertenencia a la misma clase según los dos canales de información previos.

El coeficiente de regularización es de vital importancia en este proceso, ya que es en él donde radica realmente el hecho de la combinación de clasificadores. En efecto, en su cómputo intervienen por un lado los valores de los estados de los nodos a través de $\mu_i^j(t)$, cuyos valores iniciales son los proporcionados como salidas por el clasificador individual FC en su proceso de decisión, tal y como se ha mencionado previamente; por otro lado, también intervienen las probabilidades proporcionadas por BP tal y como se acaba de describir más arriba. Esto expresa la existencia de un proceso de combinación FC y BP mediante dicho coeficiente.

b) Definición del coeficiente contextual

En cada iteración t , incluso desde el instante inicial cuando t es igual a cero, el nodo i tiene asignado un valor de estado en cada red_j dado por $\mu_i^j(t)$. Pues bien, según los valores de los diferentes estados en las distintas redes, dicho nodo se puede clasificar mediante la ecuación (4.20), que es en realidad una adaptación de la regla de decisión dada en la ecuación (3.7) para grados de pertenencia bajo la filosofía del clasificador FC individual.

$$i \in w_j \text{ si } \mu_i^j(t) > \mu_i^h(t) \quad \forall h \neq j \text{ con } j, h = 1, 2, \dots, c \quad (4.20)$$

Bajo el mismo concepto de vecindad, expresado previamente como N_i^m , aplicando la regla de decisión (4.20), el nodo (píxel) i habrá sido clasificado como perteneciente a una clase dada, supóngase w_f , mientras que el nodo (píxel) $k \in N_i^m$ se habrá clasificado perteneciendo a la clase w_g , siendo $f, g = 1, \dots, c$. En cuyo caso, al nodo i se le asigna la etiqueta $l_i = f$ y al nodo k la etiqueta $l_k = g$. Pues bien, el coeficiente *contextual* en la iteración t se define teniendo en cuenta las etiquetas anteriores tal y como se expresa a continuación,

$$c_{ik}(t) = \begin{cases} +1 & l_i = l_k & k \in N_i^m, \quad i \neq k \\ -1 & l_i \neq l_k & k \in N_i^m, \quad i \neq k \\ 0 & & k \notin N_i^m, \quad i = k \end{cases} \quad (4.21)$$

Según la ecuación (4.21) el coeficiente contextual expresa que si el nodo i y su vecino k , siendo ambos distintos, pertenecen a la misma clase, entonces entre ellos existe lo que se denomina una compatibilidad contextual, en cuyo caso, el valor de dicho coeficiente toma el máximo valor de +1. Por el contrario, cuando ambos nodos pertenecen a clases diferentes entonces se habla de incompatibilidad contextual y el valor de dicho coeficiente toma el mínimo valor de -1. Obsérvese cómo el coeficiente c_{ik} no depende del índice j , lo que equivale a decir que es independiente de la red_j y por extensión de la clase w_j .

c) Combinación de los coeficientes de regularización y contextual

Cuando existe compatibilidad entre el nodo i y sus vecinos k , el valor de su estado varía en una mínima cantidad; se dice en este caso que el nodo i está próximo a la estabilidad.

Por el contrario, si existen incompatibilidades entre ellos, el estado del nodo i se modificará sustancialmente según la magnitud del grado de incompatibilidad, aunque a medida que el proceso iterativo progresa, el nodo i también tenderá hacia la estabilidad.

Una vez definidos los dos coeficientes, regularización y contextual, el objetivo se centra ahora en obtener los pesos causales de conexión $s_{ik}^j(t)$ que aparecen en la ecuación (4.17) a partir de dichos coeficientes. Esto se lleva a cabo mediante la combinación apropiada de ambos coeficientes, esta combinación se puede realizar de diferentes maneras, habiendo optado en este trabajo de investigación por una media ponderada como sigue,

$$W_{ik}^j(t) = \rho r_{ik}^j(t) + (1 - \rho) c_{ik}(t); \quad s_{ik}^j = \left[\text{sgn}(W_{ik}^j) \right]^v W_{ik}^j; \quad \text{sgn}(W_{ik}^j) = \begin{cases} -1 & W_{ik}^j \leq 0 \\ +1 & W_{ik}^j > 0 \end{cases} \quad (4.22)$$

donde $\rho \in [0,1]$ determina la importancia concedida a ambos coeficientes. En nuestros experimentos se ha optado por elegir el valor de $\rho = 0.80$. El hecho de conceder más importancia al coeficiente de regularización que al contextual estriba en la evidencia de que $c_{ik}(t)$ contribuye sólo con valores discretos de $+1$ ó -1 según se trate de una recompensa o una penalización, mientras que los valores de $r_{ik}^j(t)$ pertenecen al rango $[-1, +1]$. En consecuencia las contribuciones del coeficiente contextual son más bruscas que las del de regularización y por tanto, se ha considerado conveniente reducir la influencia de aquél. Se han probado diferentes valores de ρ habiéndose llegado a la conclusión de que dicho valor es adecuado para la aplicación desarrollada. En la ecuación (4.22), sgn representa la función matemática *signo* y v es el número de valores negativos en el conjunto $C \equiv \{W_{ik}^j(t), r_{ik}^j(t), c_{ik}(t)\}$; una definición formal para v es la siguiente: $S \equiv \{q \in C / q < 0\} \subseteq C$, $v = \text{card}(S)$.

Para ilustrar las ideas expuestas hasta aquí, sobre la estrategia de combinación mediante el método basado en el paradigma de los *MCFs*, en la figura 4.4 se muestra un esquema gráfico para un caso concreto con cuatro clases. En el nivel de iteración t se observa la existencia de cuatro redes, asociadas con las cuatro clases disponibles. Los valores de los estados de los nodos se designan por μ_i^j para el nodo central y μ_k^j para los k vecinos, entendiendo que se refieren en ambos casos a la iteración t . Paralelamente, se dispone

de las probabilidades proporcionadas como salidas por el clasificador individual BP durante su fase de decisión, tanto para el nodo i como para sus k vecinos, permaneciendo invariables durante todo el proceso de actualización. En el instante correspondiente a la iteración t se clasifican los nodos i y k como pertenecientes a una de las clases disponibles de acuerdo con los valores de sus estados en dicho instante. Dicha clasificación se lleva a cabo mediante la regla de decisión dada en (4.20). Tras la clasificación se etiquetan con l_i y l_k respectivamente, según su pertenencia a una de las clases, tal y como se ha descrito previamente. En el instante $t + 1$ se obtienen los correspondientes valores de los coeficientes de regularización y contextual según se ha explicado anteriormente. El proceso continúa de forma iterativa siguiendo este mismo esquema.

4.4.3.5 Definición del factor de decadencia

La introducción del factor de decadencia en el proceso de actualización se basa en la hipótesis de que para garantizar una estabilidad relativamente alta en el estado de un nodo i en la red_j , se podría tratar de desestabilizarlo en una pequeña cantidad en la iteración t con el fin de asegurarnos de que si es verdaderamente estable, a pesar de esta pequeña desestabilización el nodo debe tender de nuevo hacia la estabilidad. Con esto se pretende garantizar, de alguna manera, que el proceso de actualización tiene la fuerza suficiente como para llevar los nodos, y por extensión las redes, hacia estados estables.

La cuestión que se plantea ahora consiste en determinar qué criterio aplicar para definir el factor de decadencia. La idea subyacente consiste en determinar la variación relativa del estado del nodo i frente a las variaciones producidas en los valores de los estados de los nodos vecinos.

Con el propósito expresado previamente y con el fin de definir dicho factor, se construye un acumulador de celdas de dimensión $n = M \times N$ donde cada celda i se asocia con el nodo de idéntico nombre. Obsérvese, cómo el número de celdas coincide con el tamaño de la imagen que se está clasificando. La celda i contiene el número de veces, h_i^j , que el nodo i ha cambiado el valor de su estado de forma significativa en la red_j . Todos los valores h_i^j se inicializan a cero. Posteriormente, en cada iteración se verifica si $|\mu_i^j(t+1) - \mu_i^j(t)| > \varepsilon$, en cuyo caso se incrementa el valor acumulado en la celda i en

una unidad, esto es $h_i^j = h_i^j + 1$; ε es el valor que determina si el cambio se considera o no significativo. En los experimentos llevados a cabo en este trabajo se ha fijado a 0.05 considerándose un valor aceptable, ya que valores inferiores tienden a aumentar el número de iteraciones sin mejoras significativas en los resultados.

La estabilidad del nodo i se mide como la fracción de cambios acumulados por la celda i frente a los cambios producidos en su vecindad $k \in N_i^m$ y el número de iteraciones t . Finalmente, el factor de decadencia se determina como sigue,

$$d_i^j = \begin{cases} 0 & h_i^j = 0 \text{ y } h_k^j = 0 \\ \frac{h_i^j}{(\bar{h}_k^j + h_i^j)t} & \text{de otro modo} \end{cases} \quad (4.23)$$

donde h_i^j se define como se ha mencionado anteriormente y \bar{h}_k^j es el número medio de cambios acumulados por los nodos k en la vecindad N_i^8 . Tal y como se deduce de la ecuación (4.23), si $h_i^j = 0$ y $h_k^j = 0$ el factor de decadencia toma el valor nulo, esto significa que no se producen cambios significativos en el estado i de la red_j y por tanto se asume que dicho nodo presenta una estabilidad alta; si la fracción de cambios, relativa a la vecindad definida por $h_i^j / (\bar{h}_k^j + h_i^j)$, es pequeña, la estabilidad del nodo i es también alta, cumpliéndose en este caso que d_i^j tiende a cero, lo que expresa que el factor de decadencia apenas influirá en la desestabilización del nodo i debido a su ya alta estabilidad. Incluso, ante cualquier valor de la fracción de cambios, el término d_i^j disminuiría a medida que aumenta el número de iteraciones t , esto trata de reflejar la idea de que según aumenta el número de iteraciones, los nodos tienden a estabilizarse por la propia naturaleza del proceso iterativo.

Evidentemente, cuanto mayor sea el valor de d_i^j tanto mayor será el grado de desestabilización introducido en el nodo i para la red_j .

4.4.3.6 Síntesis del proceso global basado en los MCFs

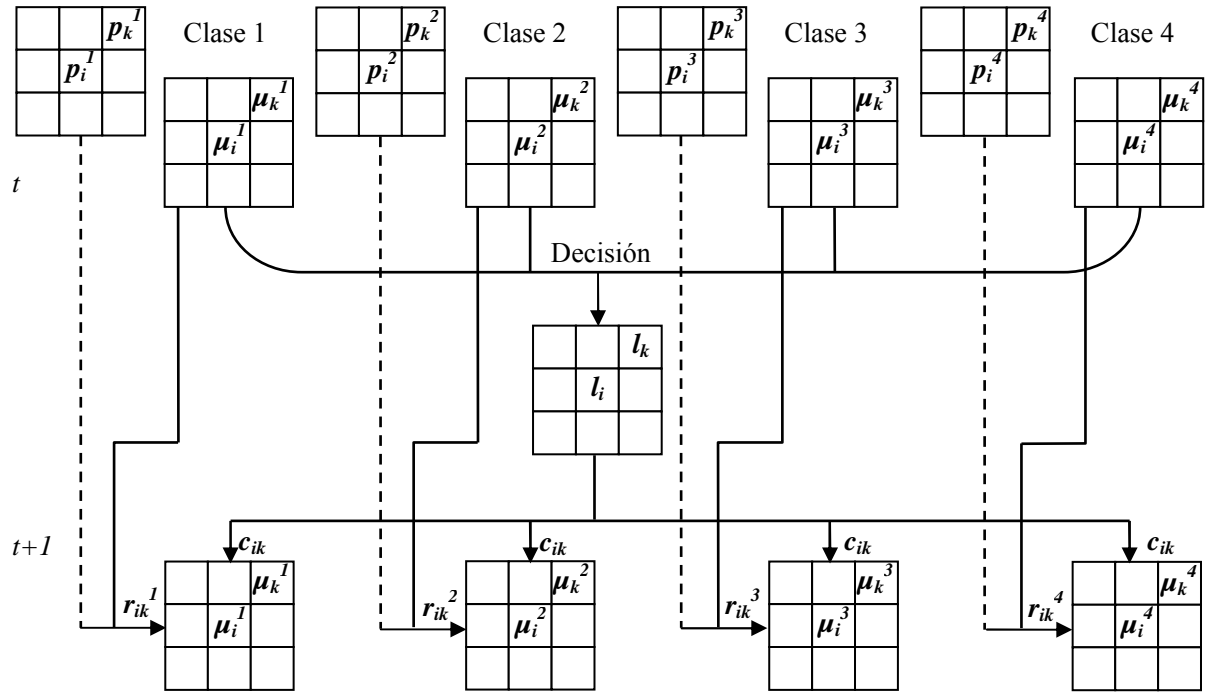
En esta sección se sintetiza el proceso global de combinación de clasificadores mediante el paradigma de los *MCFs*; si bien, antes es necesario abordar una cuestión previa

derivada de su naturaleza iterativa, ésta es la relacionada con los criterios según los cuales se detiene el proceso.

Uno de tales criterios ha de ser necesariamente el de *convergencia*, que se determina en función del grado de estabilidad de cada nodo en particular y de cada una de las redes en general. La idea intuitiva sobre la convergencia ha sido introducida en la sección previa. Así, se dice que el nodo i es estable si el valor del estado entre dos iteraciones consecutivas no cambia significativamente, entendiendo por un cambio significativo exactamente lo mismo que durante la definición del factor de decadencia. En definitiva, el proceso iterativo se dice que *converge* si todos los nodos en todas las redes cumplen la siguiente relación: $|\mu_i^j(t-1) - \mu_i^j(t)| > \varepsilon$. Según este criterio, el máximo grado de estabilidad se lograría con un valor de tolerancia nulo, es decir para $\varepsilon = 0$. El segundo criterio surge ante la posibilidad de que no se llegue a la convergencia con el criterio anterior, por ello se establece un número máximo de iteraciones t_{max} ; si bien, en caso de detener el proceso por alcanzar dicho máximo, no se puede garantizar en absoluto la convergencia, ni siquiera el grado de aproximación hacia ella.

El proceso *MCF* global se sintetiza como sigue:

1. *Inicialización*: se carga cada nodo en cada *red_j* con su grado de pertenencia a la clase w_j de acuerdo con la salida proporcionada por el clasificador individual *FC* en su fase de decisión, esto es con $\mu_i^j(t)$ siendo $t = 0$, según la ecuación (3.5); se establece la tolerancia a un cierto valor, en este caso $\varepsilon = 0.02$; se define el número máximo de iteraciones, en nuestros experimentos $t_{max} = 20$, sección 5.3.2.1. Finalmente, se define nc como el número de nodos que cambian sus valores de estado en cada iteración.
2. *Proceso MCF*: $t = 0$
 mientras $t < t_{max}$ ó $nc \neq 0$
 $t = t + 1$; $nc = 0$;
 para cada nodo i
 se actualiza $\mu_i^j(t)$ de acuerdo con la ecuación (4.17)
 si $|\mu_i^j(t) - \mu_i^j(t-1)| > \varepsilon$ entonces $nc = nc + 1$; *fin*
 fin;
 fin;
3. *Salidas*: los estados $\mu_i^j(t)$ actualizados para todos los nodos.
4. *Decisión*: el nodo i con su vector de atributos x_i se clasifica según la regla (4.20).


 Figura 4.4 Esquema gráfico del método de combinación basado en el esquema *MCF* para cuatro clases

4.5 Combinación de clasificadores por relajación

El procedimiento de combinación de clasificadores por *relajación* (*REL*), propuesto en este trabajo de investigación, tiene como precursor, a su vez, el trabajo de Guijarro y col. (2008). En este caso se utiliza un esquema similar, en su concepción, al propuesto en la sección anterior para los *MCFs*, si bien con las tres diferencias fundamentales siguientes:

- Se utilizan las probabilidades como los estados iniciales de los nodos, que son las salidas proporcionadas por el clasificador individual *BP* durante su fase de decisión. Por tanto, en este caso, se cambian probabilidades por grados de pertenencia, que son los utilizados en la inicialización de los *MCFs*.
- Posteriormente, son esas probabilidades las que se actualizan de acuerdo con el proceso de relajación probabilista definido en la ecuación (4.24), el cual está basado en los trabajos de investigación de Rosenfeld y col. (1976) y Hummel y Zucker (1983); sustituyendo así, al procedimiento definido en la ecuación (4.17) para los *MCFs*.

- c) Los valores de los estados de los nodos comienzan siendo probabilidades en lugar de grados de pertenencia, estas probabilidades son las que se actualizan durante el proceso iterativo, constituyendo los valores de los estados de los nodos en las sucesivas iteraciones.

En resumen, el esquema correspondiente a la fase de *entrenamiento*, es exactamente el mismo que el propuesto en la figura 4.2; mientras que las diferencias se producen en la fase de *decisión*, según la descripción que se proporciona a continuación.

Para la mencionada fase de decisión, se parte de la misma estructura de redes que en el caso de los *MCFs*, es decir una red_j por cada clase w_j . El número de nodos de cada red coincide con el número de píxeles en la imagen a clasificar. Cada nodo i en cada red_j se carga inicialmente con los valores de probabilidad proporcionados por el clasificador individual *BP*, tal y como se ha mencionado previamente. Así al nodo i en la red_j , representado por el vector \mathbf{x}_i cuyas componentes son los valores espectrales RGB del píxel asociado, se le asigna la probabilidad de pertenencia a la clase w_j dada por la ecuación (3.9), de forma que el valor del estado inicial viene dado por $p_i^j = P_b(w_j | \mathbf{x}_i)$, donde el segundo término de esta expresión resulta ser dicha probabilidad, según se describe en la sección 3.3.2.2. Como en el caso de los *MCFs* los valores de dichos estados se escalan linealmente al rango $[-1, +1]$ en lugar de su rango natural de valores de las probabilidades, esto es $[0, +1]$.

El proceso de relajación probabilista o actualización de los estados viene dado por la siguiente expresión,

$$p_i^j(t+1) = \frac{1}{2} \left[f(D_i^j) + p_i^j(t) \right] \quad (4.24)$$

donde $D_i^j = \sum_{k=1}^n s_{ik}^j(t) p_k^j(t)$ tiene el mismo significado que en el caso de la ecuación (4.17), es decir, representa la influencia externa ejercida por el resto de los nodos en la red. Como allí, mediante la aplicación de los principios de la Gestalt, dicha influencia es ejercida exclusivamente por los nodos k en la vecindad N_i^8 de i . La diferencia con respecto a la mencionada ecuación (4.17) estriba en el hecho de que ahora son las probabilidades las que se actualizan en función de la iteración t . A la vista de la ecuación (4.24), la actualización del valor del estado del nodo i , $p_i^j(t+1)$, resulta ser el

valor medio entre $f(D_i^j)$ y el propio valor de probabilidad $p_i^j(t)$ en la iteración actual. La función f se define aquí también como la función tangente hiperbólica con el mismo factor de suavizado, es decir $f(D_i^j) = \tanh(\sigma D_i^j)$ con $\sigma = 0.8$. Con esta definición de la función f se consigue que el recorrido de la misma quede limitado al intervalo $[-1, +1]$.

El coeficiente $s_{ik}^j(t)$ se define exactamente igual que en la ecuación (4.22), obtenido a partir de los denominados coeficientes de regularización y contextual. Este último definido exactamente igual que en la ecuación (4.21). La diferencia estriba en el cálculo del coeficiente de regularización con respecto al definido en la ecuación (4.19). En efecto, ahora se obtiene a partir de las probabilidades actualizadas $p_k^j(t)$ en cada iteración t , estableciéndose la compatibilidad entre los nodos i y k a través de las diferencias entre estas probabilidades y los grados de pertenencia μ_k^j , éstos últimos proporcionados como salidas por el clasificador individual FC , durante su fase de decisión. Los grados de pertenencia permanecen invariables a lo largo de todo el proceso de relajación. En consecuencia, el coeficiente de regularización se define como sigue,

$$r_{ik}^j(t) = \begin{cases} 1 - |p_i^j(t) - \mu_k^j| & k \in N_i^m, \quad i \neq k \\ 0 & k \notin N_i^m, \quad i = k \end{cases} \quad (4.25)$$

En definitiva, en este modelo, se intercambian los roles entre los clasificadores FC y BP en el sentido de que las inicializaciones y los estados varían según el modelo MCF o REL . En cualquier caso, la combinación de los clasificadores se realiza, en ambos casos, durante la fase de decisión.

Es precisamente en la fase de decisión donde finalmente se determina a qué clase pertenece el píxel i , representado por su nodo equivalente. Esto se lleva a cabo a través de los valores finales de los estados de los nodos según la regla dada en (4.26), siendo de la misma naturaleza que la definida en la ecuación (3.21).

$$i \in w_j \quad \text{si} \quad p_i^j(t) > p_i^h(t) \quad \forall h \neq j \quad \text{con} \quad j, h = 1, 2, \dots, c \quad (4.26)$$

4.6 Resumen del capítulo

Este capítulo constituye la clave del trabajo de investigación llevado a cabo. En efecto, en él se describen de forma pormenorizada las estrategias de combinación propuestas.

Dichas estrategias se enfocan tanto desde el punto de vista de la combinación píxel a píxel como teniendo en cuenta los píxeles a clasificar y sus vecinos. El primer enfoque da lugar, a su vez, a dos estrategias de índole local, una basada en el paradigma de la teoría de la decisión multi-criterio en el contexto fuzzy, identificada como *MCDMF* y la otra en la integral fuzzy, *IF*. En ellas intervienen seis clasificadores individuales *FC*, *BP*, *BN*, *SO*, *L1* y *L2*, que constituyen la base de las combinaciones. El segundo enfoque origina otras dos estrategias de índole global, una basada en el paradigma de los mapas cognitivos fuzzy *MCFs* y otra en los procesos de relajación probabilista, *REL*. En el caso de las estrategias globales se utilizan solamente dos clasificadores individuales, *FC* y *BP*. Las dos estrategias globales tienen su fundamento en sendos procesos iterativos basados en la actualización de los estados de un conjunto de nodos, que se asocian con los píxeles de las imágenes a clasificar.

Entrando más detalladamente en los métodos propuestos, en las estrategias de índole local, se ha planteado un procedimiento para determinar la importancia relativa de los clasificadores individuales con el fin de aprovechar dicha información durante la combinación de los mismos.

Por lo que respecta a las estrategias globales, se han definido dos coeficientes, llamados de regularización y contextual, que establecen ciertas relaciones de compatibilidad entre el píxel que se está clasificando y sus vecinos. Estas relaciones tienen su inspiración en los principios de la Gestalt, ya que su aplicación en el campo de las imágenes resulta relativamente fácil a la vez que intuitivo.

En todos los métodos propuestos, tanto locales como globales, la combinación se lleva a cabo durante la fase de decisión, siendo aquí donde se determina la pertenencia de cada píxel a una de las clases disponibles y establecidas durante la fase de entrenamiento.

En la fase de aprendizaje, se ha diseñado un mecanismo por el cual, a partir de un clasificador de naturaleza supervisada se consigue un procedimiento de aprendizaje no

supervisado. Esto representa un cierto avance hacia la automatización del proceso conjunto entrenamiento/decisión.

La eficacia de los métodos propuestos se verifica en el capítulo cinco, donde se diseña una batería de pruebas con tal propósito.

Capítulo 5

Análisis de resultados

5.1 Introducción

Una vez establecidos los fundamentos relativos a las estrategias combinadas propuestas en este trabajo de investigación, el objetivo del presente capítulo consiste en verificar y validar las mismas. Como se ha mencionado reiteradamente, dichas estrategias constan de una fase de entrenamiento y una fase de decisión. La fase de entrenamiento se ha diseñado de forma tal que a partir de métodos de clasificación de naturaleza supervisada se propone un esquema de aprendizaje *no supervisado*. En la fase de decisión es donde realmente se lleva a cabo la *combinación* de los clasificadores que intervienen en cada uno de los métodos propuestos.

Las estrategias combinadas se han aplicado a imágenes aéreas con texturas naturales. Tanto los métodos individuales como las estrategias combinadas, en el primer caso por su naturaleza supervisada, requieren el ajuste de ciertos parámetros bien en la fase de aprendizaje o bien en la de decisión. Con el fin de proceder a dicho ajuste, se han utilizado dos tipos de datos: *a)* muestras procedentes de bases de datos relativamente bien conocidas en el ámbito de los clasificadores, no necesariamente imágenes y *b)* muestras procedentes de diversas imágenes con texturas naturales.

Dada la naturaleza no supervisada del diseño, se requiere estimar la mejor partición posible, es decir el mejor número de clases posibles para las imágenes que van a ser utilizadas en los experimentos y de las que no se conoce dicho número. Esta estimación se realiza durante el proceso de entrenamiento. Tras lo cual, se puede determinar el comportamiento de los clasificadores individuales y a partir de ahí deducir sus

competencias relativas, que son utilizadas en las estrategias combinadas locales propuestas en el capítulo cuatro.

Por otro lado, al tratarse de procedimientos con aprendizaje involucrado, se ha diseñado una estrategia de test con el fin de verificar el comportamiento de los diferentes métodos, a medida que el sistema adquiere un mayor grado de conocimiento por el entrenamiento.

Finalmente, se comparan los resultados obtenidos mediante las estrategias combinadas propuestas, tanto frente a los métodos de clasificación individuales como frente a otras estrategias combinadas clásicas tales como las descritas en la sección 3.5.2. Para establecer dichas comparaciones, es necesario disponer de las imágenes de referencia. En el caso de las imágenes aéreas utilizadas en la investigación, dicha base no está disponible por su inexistencia, por lo que se ha diseñado un procedimiento para su creación.

Por consiguiente, según lo anteriormente expuesto, el capítulo se estructura de la siguiente manera:

- En la sección 5.2 se realiza una descripción de los datos utilizados, de los cuales una parte de ellos sirve para ajustar los parámetros involucrados tanto en la fase de entrenamiento como en la de decisión. Otra parte, constituida por imágenes aéreas, se utiliza para el entrenamiento y clasificación, tras el ajuste de los parámetros mencionados previamente.
- En la sección 5.3, se describen los procedimientos para el ajuste de los parámetros mencionados.
- En la sección 5.4, se describe la estrategia general de pruebas planteada, donde se realiza un proceso incremental de entrenamiento, que permitirá determinar la repercusión del grado de aprendizaje adquirido por el sistema sobre los resultados obtenidos.
- En la sección 5.5, a partir de un conjunto dado de datos con el número de clases desconocido, se establece el procedimiento para determinar cuál es el mejor número de clases y con ello la mejor partición del conjunto. Esto constituye el

fundamento de las estrategias no supervisadas propuestas en el capítulo cuatro. Mediante el mismo procedimiento, se obtienen las competencias de cada clasificador individual y a partir de ahí los pesos específicos y las densidades fuzzy, que se utilizan respectivamente en las estrategias combinadas *MCDMF* e *IF*.

- En la sección 5.6, se describe, el procedimiento establecido para generar las imágenes de referencia con las que verificar el comportamiento de los diferentes métodos de clasificación y de esta forma poder validar las estrategias propuestas en este trabajo de investigación.
- Finalmente, en la sección 5.7, se recogen los resultados obtenidos con todos los clasificadores previstos para las imágenes utilizadas en los experimentos. A partir de dichos resultados se realiza el correspondiente análisis comparativo y se extraen las conclusiones pertinentes.

5.2 Descripción de los datos utilizados

Existen dos tipos de datos claramente diferenciados, a saber:

- a) Datos genéricos utilizados para ajustar los parámetros requeridos por los métodos, bien durante la fase de entrenamiento o bien durante la de decisión.
- b) Imágenes aéreas de texturas naturales sobre las que se aplican los procesos de entrenamiento y decisión en sus respectivas fases. Se trata de las imágenes con las que se obtienen los resultados de clasificación tanto con los clasificadores descritos en el capítulo tres como los propuestos en el cuatro, permitiendo así validar el comportamiento de estos últimos.

5.2.1 Descripción de los datos genéricos

En la tabla 5.1 se describen brevemente las 16 categorías de datos utilizadas con el propósito de ajustar los parámetros involucrados durante la fase de entrenamiento según se describe posteriormente. El calificativo de genéricos, hace referencia a su distinta naturaleza y procedencia.

El conjunto de datos numerados del 1 al 9 proceden del conocido “Machine Learning Repository” (Asunción y Newman, 2008). Por tanto, las características de estos datos vienen determinadas por las especificaciones dadas en el mencionado repositorio. En la columna descripción de dicha tabla, se proporciona un resumen sobre el contenido de cada conjunto.

Las categorías numeradas del 10 al 12 corresponden a datos generados de forma sintética, es decir, distribuyendo los datos manualmente según las características especificadas en la columna descripción.

Las categorías 13 a 16 son datos procedentes de imágenes de exterior capturadas bajo las condiciones especificadas en la columna descripción. Estos datos son exactamente vectores x cuyas componentes son los valores espectrales de los píxeles en el modelo de color RGB. En síntesis, proceden de imágenes obtenidas bajo diferentes condiciones de iluminación correspondiendo a días soleados y días con alta nubosidad. En la figura 5.1 (a) y (b) se muestran dos imágenes de exterior correspondientes a estas cuatro últimas categorías. En (a) aparece una imagen tomada en un día soleado y en (b) otra imagen correspondiente a un día con una elevada nubosidad. En ambos casos, han sido capturadas mediante un sistema de lentes conocidas como de “ojo de pez”. Estas imágenes son cortesía del Centro de Investigación Forestal (CIFOR) del Instituto Nacional de Tecnología Agraria y Alimentaria (INIA) perteneciente al Ministerio de Ciencia e Innovación.

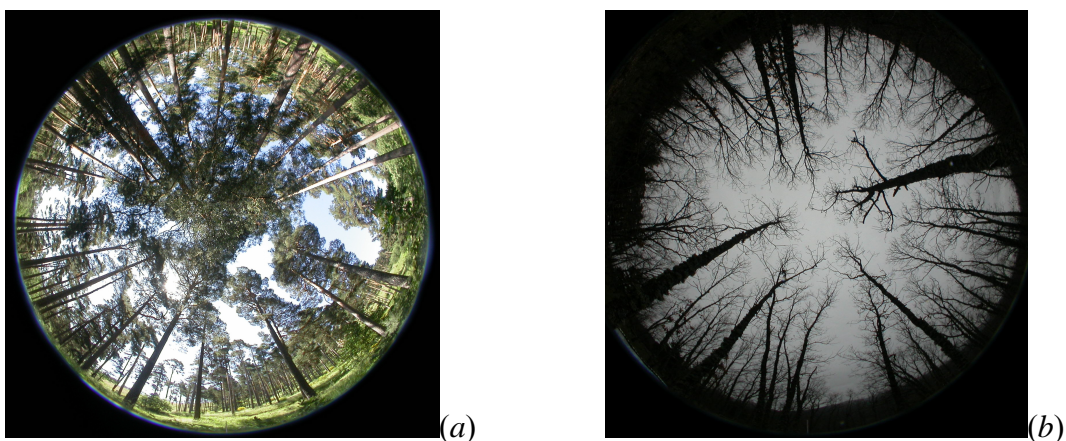


Figura 5.1 Imágenes hemisféricas de zonas boscosas (a) tomada en un día soleado; (b) tomada en un día nublado

En la columna identificada como q , en la tabla 5.1, se muestra la dimensionalidad de los datos; es decir, el espacio q -dimensional al que pertenecen, de suerte que dado un dato x

de una determinada categoría, $\mathbf{x} \in \mathcal{R}^q$; en la columna c aparece el número de clases en las que se distribuyen los datos, dicho número está fijado de antemano, esto significa que no se determinan mediante las estrategias no supervisadas descritas en este trabajo; en la columna n se proporciona el número de muestras disponibles para cada categoría.

Tabla 5.1 Descripción de los conjuntos de datos utilizados para el ajuste de parámetros en la fase de entrenamiento

#	Categorías de datos	q	c	n	Descripción
1	Bupa liver disorders	6	2	345	BUPA Medical Research Ltd.: resultados de análisis de sangre, cuya presencia parece que sean los causantes de desórdenes en el funcionamiento del hígado.
2	Cloud	10	2	2048	P. Collard del California Space Institute: canales espectrales infrarrojo y visible para nubes en imágenes de satélite.
3	Glass	10	6	214	B. German, V. Spiehler from Central Research Establishment & Diagnostic Products Corporation: diferentes propiedades de cristales.
4	ImageSegm	19	7	2310	Vision Group, University of Massachusetts & C. Brodley: clasificación de imágenes segmentadas manualmente.
5	Iris	4	3	150	R.A. Fisher & M. Marshall: diferentes clases de lirios (plantas).
6	Magi4	11	2	19020	R. K. Bock Major Atmospheric Gamma Imaging Cherenkov Telescope project (MAGIC) & P. Savicky Institute of Computer Science: partículas gamma de alta energía en un sustrato atmosférico obtenidas mediante el telescopio gamma Cherenkov.
7	Thyroid	5	3	215	Ross Quinlan Danny Coomans & Stefan Aeberhard: datos de la glándula tiroidea.
8	PimaIndians	8	2	768	National Institute of Diabetes and Digestive and Kidney Diseases (V. Sigillito): parámetros de herencia en la tribu India Pima.
9	Wine	13	3	178	Forina, M. et al, PARVUS & S. Aeberhard: análisis químicos de distintas clases de vinos.
10	Sintéticas 1	1	2	132	Datos generados manualmente y distribuidos en dos clases muy bien separadas.
11	Sintéticas 2	2	3	164	Datos generados manualmente y distribuidos en tres clases con solapamiento.
12	Sintéticas 3	3	4	152	Datos generados manualmente y distribuidos en dos clases que se solapan y en otras dos perfectamente separadas.
13	Imágenes de exterior 1	3	2	1270	Datos extraídos manualmente a partir de una imagen procedente de un bosque de pinos, tomada en un día soleado.
14	Imágenes de exterior 2	3	3	1434	Datos extraídos manualmente a partir de una imagen procedente de un bosque de robles, tomada en un día soleado.
15	Imágenes de exterior 3	3	4	4200	Datos extraídos manualmente a partir de una imagen procedente de un bosque de pinos, tomada en un día nublado.
16	Imágenes de exterior 4	3	5	4500	Datos extraídos manualmente a partir de una imagen procedente de un bosque de robles, tomada en un día nublado.

5.2.2 Descripción de las imágenes aéreas

Se dispone de un conjunto de 52 imágenes aéreas adquiridas durante el mes de Mayo del año 2006 en la región de Abadín en la provincia de Lugo (España). Dichas imágenes han sido proporcionadas por la empresa Dimap (2008) como parte del material de trabajo para el Servicio Territorial de Galicia (SITGA). Aunque existe un mayor número de imágenes disponibles, se han seleccionado las anteriores por constituir un conjunto significativo y a la vez suficientemente descriptivo según los criterios de los expertos en la compañía Dimap para el propósito previsto.

Uno de los objetivos a conseguir por Dimap, respecto del análisis de dichas imágenes, se centra en la identificación de las texturas naturales existentes en las mismas con la mayor precisión posible. Su aplicación final se orienta hacia tareas relativas a la ordenación del territorio y de prevención sobre posibles catástrofes naturales, tales como incendios.

a) Descripción específica de las imágenes

A continuación se describen una serie de características inherentes a este conjunto de imágenes, dada su importancia en nuestros experimentos.

1. En su origen, dichas imágenes se encuentran georreferenciadas en coordenadas geográficas UTM (Universal Transverse Mercator), donde las medidas se establecen en metros. El número de zona UTM se corresponde con la 29, siendo T la banda, situándose por tanto, en una latitud Norte comprendida entre 40° y 48°.
2. La superficie geográfica fotografiada posee una extensión de $552,26 \text{ Km}^2$, según los datos de captura proporcionados por el suministrador.
3. Su captura ha sido realizada mediante vuelos aéreos a distintas alturas de vuelo, variando entre 820 m y 1040 m. Este hecho hace que la equivalencia entre coordenadas geográficas y coordenadas imagen varíen ligeramente según las diferentes imágenes.
4. Las tomas se han realizado bajo proyección ortogonal sobre la superficie terrestre, es decir el eje óptico del sistema de captura se posiciona perpendicularmente a la dirección de vuelo.

5. Las imágenes utilizadas en nuestros experimentos se han formateado convenientemente, de forma que su resolución final resulta ser de 512×512 píxeles en las dimensiones x e y , según el sistema de coordenadas de la imagen. La resolución espacial es de aproximadamente 1 m/píxel tanto en la dirección x como y .
6. Cada píxel se caracteriza por sus tres componentes del espectro visible en el modelo de color RGB (Pajares y Cruz, 2007).

Desde el punto de vista de la identificación de las texturas naturales bajo el planteamiento llevado a cabo en este trabajo, no es necesario tener en cuenta ni las coordenadas geográficas ni la información de las alturas del terreno proporcionadas por los Mapas Digitales del Terreno (MDT), que también se encuentran disponibles conjuntamente con las imágenes. Por este motivo, para la clasificación de las texturas naturales mediante los métodos de clasificación utilizados en la investigación, se han obviado tanto la georreferenciación como la información de alturas del terreno. Si bien, en futuros trabajos podría considerarse la incorporación de ambos, con el fin de discernir mejor entre diferentes clases de texturas. En efecto, según la ubicación geográfica y las alturas del terreno, la vegetación varía considerablemente y por consiguiente, a partir de este conocimiento determinadas ambigüedades podrían eliminarse bajo la consideración de la mencionada información.

b) Distribución de las imágenes para las pruebas

De las 52 imágenes disponibles, se eligen 12 aleatoriamente con el fin de seleccionar sobre las mismas el conjunto inicial de muestras de entrenamiento con el que se inician los procesos de entrenamiento según los métodos individuales, como se verá seguidamente. Este conjunto se denomina SI. Dado que se trata de un diseño basado en una estrategia de naturaleza *no supervisada*, las texturas involucradas en estas imágenes se encuentran sin clasificar. Por tanto, se dispone de las muestras correspondientes sin conocer ni el número de clases apropiado para ellas, ni por supuesto su distribución en las clases.

Las 40 imágenes restantes se distribuyen equitativamente y de forma aleatoria en cuatro conjuntos, que designamos como S0, S1, S2 y S3. Cada uno de estos conjuntos contiene 10 imágenes.

La distribución de las imágenes en los conjuntos descritos anteriormente está motivada por el planteamiento de la estrategia de pruebas propuesta, que se describe posteriormente en la sección 5.4. Como anticipo, las imágenes pertenecientes al conjunto SI se utilizan para determinar el número de clases a partir de la estrategia diseñada de aprendizaje no supervisado. El conjunto SI además sirve para el ajuste de ciertos parámetros involucrados en el proceso, que son necesarios establecer o estimar de antemano. Las imágenes del conjunto S0, se utilizarán como testigo para determinar cómo influye en el resultado el grado de aprendizaje adquirido por el sistema tras las sucesivas fases de entrenamiento. Finalmente, las imágenes pertenecientes a los conjuntos S1, S2 y S3 se utilizan como imágenes a clasificar de acuerdo con los tres pasos de prueba establecidos a tal fin, según se describe en la sección 5.4. Además, las muestras clasificadas, procedentes de estos últimos tres conjuntos, son utilizadas como nuevas muestras de entrenamiento con el fin de incrementar el grado de aprendizaje, tal y como se especifica también en la mencionada sección.

En la figura 5.2 se muestran dos imágenes representativas de los conjuntos SI y S0 mencionados previamente. La parte (a) corresponde a una de las imágenes perteneciente al conjunto SI, mientras que la parte (b) pertenece al conjunto S0.

Por otro lado, en la figura 5.3 (a), (b) y (c) aparecen representadas tres imágenes pertenecientes a cada uno de los conjuntos de prueba S1, S2 y S3 respectivamente.

Mediante la simple observación visual de estas imágenes, así como de aquellas otras que no se muestran pero que se utilizan en los experimentos, se pueden apreciar varios tipos de texturas naturales, entre las que destacan:

- masas de arbustos con tonalidades verdosas y muy oscuras
- praderas y campos de cultivo con tonalidades verde claro
- campos de cultivo en barbecho con tonalidades ocres, por su carácter arcilloso
- construcciones artificiales con tonalidades variando entre el ocre, marrón y blanco
- caminos de tierra arcillosa con tendencia al ocre



(a)



(b)

Figura 5.2 Imágenes representativas de los conjuntos utilizados: (a) imagen perteneciente al grupo de entrenamiento inicial SI; (b) imagen representativa del conjunto testigo S0



(a)



(b)



(c)

Figura 5.3 Imágenes representativas de los conjuntos utilizados: (a) S1; (b) S2 y (c) S3

Un aspecto importante a destacar es el hecho de que no siempre aparecen las mismas texturas en todas las imágenes. Obsérvese a propósito, cómo la imagen 5.3(b) carece de texturas de los tipos campos de cultivo en barbecho o de construcciones artificiales, que son texturas muy abundantes en las imágenes de las figuras 5.3(a) y (c). Si bien es cierto, que en mayor o menor medida, los diferentes tipos de texturas descritos previamente aparecen en unas u otras imágenes de los conjuntos SI, S0, S1, S2 y S3. Este hecho garantiza que el sistema tenga el suficiente conocimiento sobre los diferentes tipos de texturas existentes. Por otro lado, aparecen algunas texturas que, por su naturaleza espectral, tienden a constituirse en una única clase de textura desde el punto de vista de la clasificación, tal es el caso de los caminos y las construcciones por ejemplo, que poseen en muchos casos tonalidades ocreas similares.

5.3 Ajuste de parámetros

Tanto en la fase de entrenamiento como en la de decisión, es necesario ajustar determinados parámetros que intervienen, bien en los clasificadores individuales, bien en los combinados o bien de forma general en el proceso. Este es el caso de los clasificadores simples *FC*, *BN*, *SO*, *L1* y *L2*, que necesitan ajustar ciertos parámetros durante la fase de entrenamiento o de los métodos combinados tal como los *MCFs* y el operador de Yager, que requieren un ajuste de parámetros durante la fase de decisión. Por otro lado, durante el entrenamiento, también es preciso determinar el umbral, mediante el cual se determina el número óptimo de clases para un conjunto de datos del que no se conoce dicho número. El propósito, en este último caso, es el diseño de una estrategia de naturaleza no supervisada a partir de un método supervisado. A continuación se describen los procesos necesarios para el ajuste de parámetros, distinguiendo entre las consabidas fases de entrenamiento y decisión.

5.3.1 Fase de entrenamiento

Para llevar a cabo el ajuste de los parámetros durante la fase de entrenamiento se utilizan los datos procedentes de las 16 categorías disponibles, descritas en la sección 5.2.1. Tal y como se indica allí, tanto el número de clases como la distribución de los datos en las clases resultan conocidos. Este hecho permite aplicar los correspondientes procesos de aprendizaje supervisado a los clasificadores individuales *FC*, *BN*, *SO*, *L1* y *L2* descritos en el capítulo tres, sin necesidad de estimar el número de clases óptimo.

El procedimiento elegido para estimar la mayoría de los parámetros es el conocido como *validación cruzada* o “cross-validation” en la literatura inglesa especializada (Duda y col., 2001). Para llevarlo a cabo se procede como sigue. Cada uno de los conjuntos que describen las 16 categorías de datos se divide en dos partes, cada parte se elige de forma aleatoria. Una primera parte contiene el 90% de las muestras, a este conjunto se le conoce como conjunto de *entrenamiento*, designado como E_i . La segunda parte contiene el otro 10% de las muestras, denominado conjunto de *validación* y designado como V_i . Se dispone así de dieciséis conjuntos de cada tipo. Para estimar el mejor valor de un parámetro determinado de acuerdo con cada método de clasificación, se varía el valor de dicho parámetro en un rango de valores establecido como posible o válido para ese parámetro. Después, para cada valor del parámetro, se realiza un proceso de aprendizaje con las muestras pertenecientes al conjunto de entrenamiento, tras lo cual se procede a la clasificación de las muestras pertenecientes al conjunto de validación en la correspondiente fase de decisión según el clasificador en cuestión. Los resultados así obtenidos, por el clasificador, se comparan con las clasificaciones conocidas del conjunto de validación, gracias a lo cual se puede determinar el error cometido por cada clasificador en cada conjunto de validación. Se obtiene un valor de error promedio con las dieciséis categorías de datos. Aquél valor del parámetro que produzca el menor porcentaje de error medio es el finalmente elegido.

A continuación se describen exactamente los procesos llevados a cabo para determinar cada uno de los parámetros libres requeridos en los métodos de clasificación individual descritos en el capítulo tres.

5.3.1.1 Parámetros del clasificador borroso

En la sección 3.3.1.1 se describe el proceso de entrenamiento del método *FC*, donde además se identifican los parámetros necesarios, que deben estimarse previamente. Dichos parámetros, tal y como se recogen en la tabla 3.1 son: el peso exponencial m , que aparece en la ecuación (3.2); el umbral de convergencia ε y el máximo número de iteraciones t_{max} .

Como se ha mencionado previamente, el número de clases y la distribución de las muestras en las clases son conocidos, tabla 5.1. El proceso de estimación se sintetiza como sigue:

- 1) Se fija m al valor 2.0, ya que es el valor más ampliamente utilizado en diversos experimentos descritos en la literatura.
- 2) Se varía ε desde 0.01 hasta 0.1 en pasos de 0.015. Para cada uno de estos valores se estiman los centros de las clases y los grados de pertenencia de las muestras para cada uno de los dieciséis conjuntos de entrenamiento disponibles.
- 3) Seguidamente, se calcula el porcentaje de error en la clasificación para cada uno de los dieciséis conjuntos de validación.
- 4) El máximo valor del error se obtiene para $\varepsilon = 0.1$ con 10 iteraciones, mientras que el mínimo se consigue con $\varepsilon = 0.01$ y 48 iteraciones. A título informativo, los mayores porcentajes de error se producen en las categorías trece y catorce, mientras que los menores corresponden a las categorías 10, 11 y 12. Estos resultados entrañan cierta lógica, ya que en el primer caso los datos proceden de imágenes de exterior con grandes variaciones de intensidad, mientras que en el segundo son datos sintéticos. Por tanto, se eligen los valores de los parámetros correspondientes al mínimo porcentaje de error.
- 5) Fijados de este modo ε a 0.01 y el número de iteraciones a 50, ampliando así el margen de confianza en el número de iteraciones en dos unidades, se varía m desde 1.1 hasta 4.0 en pasos de 0.1. El valor de 4.0 es el valor máximo más frecuentemente utilizado en diversos experimentos. Para cada uno de estos valores de m se vuelven a estimar los centros y grados de pertenencia de las muestras a las clases empleando de nuevo los dieciséis conjuntos de aprendizaje.
- 6) Seguidamente, se utilizan de nuevo los dieciséis conjuntos de validación, para obtener el porcentaje medio de error. De esta forma se determina que el mínimo error se obtiene precisamente para un valor de m igual a 2.1.
- 7) Por consiguiente, el ajuste final resulta ser $m = 2.1$, $\varepsilon = 0.01$ y $t_{max} = 50$.

5.3.1.2 Parámetros del estimador de la ventana de Parzen

En el caso del estimador BN , es en la ecuación (3.24) donde aparecen dos parámetros, que deben ser estimados, se trata de r y h . En la ecuación (5.1) reproducimos la (3.24) por comodidad.

$$h_j = hn_j^{-\frac{r}{q}} \quad \text{para} \quad 0 \leq r \leq 0.5 \quad (5.1)$$

donde q representa, como siempre, la dimensión de los datos según la tabla 5.1. La forma de proceder en este caso es como sigue.

Según el análisis de la sección 3.3.3.1, se conoce la relación entre h_j y r por medio de la gráfica mostrada en la figura 3.2. Además, de cada uno de los dieciséis conjuntos de entrenamiento, E_i ($i = 1, \dots, 16$) de la tabla 5.1, conocemos sus clases w_j y la distribución de las muestras en clases, por tanto los n_j resultan conocidos. Esto es, dado un subconjunto cualquiera de entrenamiento E_i se conocen $w_1(n_1), \dots, w_j(n_j), \dots, w_c(n_c)$ para las muestras de E_i . Seleccionado E_i y dado un r , obtenemos un h_j fijo según la gráfica 3.2. Con este h_j fijo, el r que lo ha generado y los valores n_j conocidos, obtenemos un valor de h a través de (5.1), que identificamos como $^j h$, esto es, $^j h = h_j n_j^{r/q}$. Finalmente, para ese valor de r calculamos h como el valor medio de los $^j h$, obteniendo así un valor de h para el r dado. Obsérvese que j se refiere a cada una de las clases w_j . De esta forma, por medio de las ecuaciones (3.23) y (3.24), estimamos las funciones de densidad de probabilidad para cada una de las clases del conjunto E_i , procediendo a clasificar las muestras del correspondiente conjunto de validación de V_i , con lo que se obtiene un porcentaje de error en la clasificación de dichas muestras. A continuación variamos r , obteniendo un nuevo h y un nuevo error de clasificación. Continuamos de este modo hasta alcanzar el límite de r igual a 0.5. Este proceso se realiza para los dieciséis conjuntos E_i y V_i disponibles. Con todas las combinaciones de valores r y h obtenidas elegimos aquella con la que se haya obtenido el menor error de clasificación, resultando ser $r = 0.3$ y $h = 1.1383$ para el conjunto de datos número 15 de la tabla 5.1.

El siguiente ejemplo trata de clarificar el proceso de estimación de h tomando como referencia el mencionado conjunto de datos número 15. En este caso, el conjunto total de muestras es de 4200, de las cuales 3780 constituyen el conjunto E_{15} de muestras de entrenamiento (90%) y el 10% restante, 420 muestras, el conjunto de validación V_{15} . Las muestras de entrenamiento se distribuyen en las cuatro clases disponibles, siendo $n_1 = 1120$, $n_2 = 846$, $n_3 = 1038$ y $n_4 = 776$. Para el valor de r igual a 0.3, se obtiene el correspondiente valor h_j , a partir de la relación establecida entre r y h_j según la gráfica de la figura 3.2, que resulta ser $h_j = 0.5743$. Pues bien, teniendo en cuenta la ecuación 5.1, se calcula un valor de h para cada una de las cuatro clases, que identificamos como $^j h$

para distinguirlos de los h_j que aparecen en dicha ecuación. Despejando h en la ecuación (5.1) e identificada como 1h , se obtiene $^1h = h_1 n_1^{r/q}$, siendo q la dimensión de los datos, que en este caso resulta ser $q = 3$, según la tabla 5.1. Por tanto, el valor obtenido para 1h con los valores anteriormente especificados es 1.1589. Procediendo del mismo modo para calcular el resto de valores $^j h$ según las otras tres clases restantes se obtienen los valores dados en la tabla 5.2. Finalmente, el valor de h se obtiene como el valor medio de los cuatro valores $^j h$, que aparece en la correspondiente columna de dicha tabla.

Tabla 5.2 Valores obtenidos para el parámetro h de la ecuación (5.1)

1h	2h	3h	4h	valor medio $h = \frac{1}{4} \sum ^j h$
1.1589	1.1269	1.1502	1.1172	1.1383

El porcentaje de error para las 420 muestras de entrenamiento ha resultado ser el menor de todos los conjuntos de datos de la tabla 5.1, en concreto el 12.2%, habiéndose obtenido, como se ha mencionado previamente, para $r = 0.3$ y $h = 1.1383$; por tanto, éstos son finalmente los valores elegidos para ambos parámetros libres utilizados en la estimación de las funciones de densidad de probabilidad del método *BN*, durante los experimentos llevados a cabo en este trabajo de investigación.

5.3.1.3 Mapas auto-organizativos

En el procedimiento SO, los parámetros a fijar son por un lado los involucrados en la función de vecindad definida en la ecuación (3.29), que reproducimos aquí por comodidad como ecuación (5.2) y por otro lado la razón de aprendizaje, $\gamma(t)$, la razón de convergencia ε y el número máximo de iteraciones t_{max} .

$$K_{\alpha(t)}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\alpha^2(t)}\right) \quad \text{con} \quad \alpha(t) = \alpha_{inicial} \left(\frac{\alpha_{final}}{\alpha_{inicial}}\right)^{t/t_{max}} \quad (5.2)$$

Tal y como se menciona en el capítulo tres $K_{\alpha(t)}(\cdot)$ debe ser tal que al principio del proceso iterativo conviene que cubra toda la vecindad. En el caso de las imágenes aéreas, que serán el objeto de la clasificación, los vectores que definen las características son los valores de las componentes espectrales R, G y B para cada píxel. Es obvio, que

la máxima distancia entre características se produce entre el valor (0, 0, 0) y el (255, 255, 255), dado que se corresponden con los vértices opuestos del cubo en el modelo de color *RGB* (Pajares y Cruz, 2007). Por tanto, si hacemos $\mathbf{x} = (0, 0, 0)$ y $\mathbf{x}' = (255, 255, 255)$ resulta $\|\mathbf{x} - \mathbf{x}'\|^2 \cong 441$ con la norma euclídea. Suponiendo que en este caso extremo la influencia del núcleo debería ser relativamente baja, se prueban los tres valores siguientes para $K_{\alpha(t)}(\mathbf{x}, \mathbf{x}')$: 10^{-3} , 10^{-4} , 10^{-5} . Con cada uno de estos valores se calcula $\alpha(t)$ mediante la primera parte de la ecuación (5.2) siendo $\|\mathbf{x} - \mathbf{x}'\|^2 \cong 441$ y $t=1$, por ser éste el valor de comienzo de las iteraciones; a continuación se hace $\alpha_{final} = \alpha_{inicial}$ y se obtiene $\alpha_{inicial} = \alpha(1)$, a través de la segunda parte de (5.2), haciendo $t_{max} = 50$ por ser éste un valor indicativo según el obtenido para el clasificador *FC*. Además se fija la razón de aprendizaje presente en la ecuación (3.33) a un valor constante, $\gamma(t) = 0.5$ y $\varepsilon = 0.01$, en este último caso, también como en el clasificador *FC*. Fijados estos valores, se procede a la clasificación de los dieciséis conjuntos de validación de la tabla 5.1, obteniendo el menor porcentaje de error para el conjunto número quince, que resulta ser del 19.1% con $K_{\alpha(t)}(\mathbf{x}, \mathbf{x}') = 10^{-5}$. Por tanto, para $t=1$ se obtiene $\alpha(1) \cong 4.38$, según la ecuación (5.2), de forma que haciendo $\alpha_{final} = \alpha_{inicial} = 4.38$ se consigue dicho valor para el núcleo de influencia.

Fijados los parámetros anteriores, se procede a variar la razón de aprendizaje desde 0.01 hasta 0.2 con incrementos de 0.005, a la vez que se varía t desde 1 hasta 50 para cada valor de dicha razón. El menor porcentaje de error se obtiene para el conjunto de validación V_{12} de la tabla 5.1, siendo del 18.8% para 42 iteraciones y una razón de aprendizaje de 0.020. Finalmente, se elige la ley de aprendizaje definida en la ecuación (5.3).

$$\gamma(t) = \frac{1}{K + t} \quad (5.3)$$

De este modo, para los valores anteriores con $t = 42$ y $\alpha(42) = 0.020$, se obtiene $K = 8$, que proporciona una función cuya pendiente varía suavemente con el número de iteraciones.

Llegados a este punto, tan sólo resta fijar el valor de ε , para lo cual se procede a variar dicho coeficiente, como se hizo previamente, desde 0.01 hasta 0.1 en pasos de 0.015, de

forma que para cada valor de ε , se varía igualmente el número de iteraciones, encontrándose finalmente que el menor porcentaje de error con los datos de validación se obtiene para el conjunto número 16 de la tabla 5.1 con un 18.6% de error para $\varepsilon = 0.01$ con un número de iteraciones de $t = 45$. De este modo los valores de los parámetros finalmente estimados son los siguientes: $\alpha_{final} = \alpha_{inicial} = 4.38$, $\varepsilon = 0.01$ y $t_{max} = 50$ con la razón de aprendizaje dada por (5.3).

5.3.1.4 Cuantización vectorial

Tanto en el caso del clasificador $L1$ como en el del $L2$, los parámetros a estimar son la razón de aprendizaje $\gamma(t)$, t_{max} y ε ; además, en el caso de $L2$ el ancho de la ventana σ , según la ecuación (3.36), que se reproduce aquí como ecuación (5.4) por comodidad,

$$\frac{d_c}{d_r} = 1 - \sigma \quad \text{y} \quad \frac{d_r}{d_c} = 1 + \sigma \quad (5.4)$$

Con tal propósito y siguiendo la filosofía expuesta previamente, que proporciona algunas pautas para establecer dichos parámetros, en primer lugar se fijan $\varepsilon = 0.01$ y $t_{max} = 50$ así como el ancho de la ventana $\sigma = 0.1$. Este ancho de ventana representa una variación del 10%, para las fracciones entre d_c y d_r o viceversa, en torno a la unidad. A continuación, como en el caso del clasificador SO , se varía la razón de aprendizaje desde 0.01 hasta 0.2 con incrementos de 0.005, a la vez que también se varía t desde 1 hasta 50 para cada conjunto de datos de entrenamiento en la tabla 5.1. Con estos valores el menor porcentaje de error se obtiene para el conjunto de datos de validación número 16 de dicha tabla con un 19.3% y 18.7% de error con $L1$ y $L2$ respectivamente para $\varepsilon = 0.01$ y con números de iteraciones de $t = 43$ en $L1$ y $t = 46$ en $L2$. Se observa la similitud de valores obtenidos aquí con los del clasificador SO ; por este motivo se establece la misma razón de aprendizaje de acuerdo con la ley dada por la ecuación (5.3). Fijada la razón de aprendizaje, se varían de nuevo ε y t como en SO , obteniendo porcentajes de error similares a los obtenidos en el caso del SO con valores de $\varepsilon = 0.01$ y $t_{max} = 50$.

Finalmente, fijados los parámetros previos, se varía el ancho de la ventana desde $\sigma = 0.0$ hasta $\sigma = 0.2$, obteniéndose en este caso el menor porcentaje de error para los datos de validación procedentes del conjunto V_{14} de la tabla 5.1, resultando ser de 18.5% con $\sigma = 0.08$, quedando finalmente fijado el ancho de la ventana a este valor.

5.3.1.5 Umbral de las particiones

En la ecuación (3.43) se establece el criterio para determinar la mejor partición dado un cierto conjunto de datos, cuya distribución en clases resulta desconocida. A partir de los conjuntos de datos descritos en la tabla 5.1 y dado que tanto el número de clases como la distribución de los mismos en las clases son conocidos, se puede aplicar la función objetivo definida en dicha ecuación (3.43), que reproducimos de nuevo por comodidad como ecuación (5.5),

$$\hat{\Delta}^f(c) = \frac{\Delta^f(c)}{\sum_c \Delta^f(c)} \quad \text{con } c = 1, \dots, G-1 \quad (5.5)$$

Ahora, el objetivo se centra en estimar para qué valores de $\hat{\Delta}^f(c)$ la partición puede considerarse como válida.

Del estudio realizado en la sección 3.4.4 se desprende que tanto la función criterio *suma del error al cuadrado* (SE) como la función criterio de *mínima varianza relacionada* (VC) tienen comportamientos similares (ver figura 3.6). Por tanto, en la ecuación (5.5) tomamos f como SE y VC ; además, como el máximo número de clases en las diferentes categorías de datos en la tabla 5.1 es $c = 7$, establecemos el valor de G igual a 8. De este modo, para cada uno de los dieciséis conjuntos de entrenamiento E_i de la tabla 5.1, se aplica el procedimiento no supervisado propuesto en la sección 4.2.1, de suerte que comenzando en c igual a 2 y hasta c igual a $G - 1$ se distribuyen las muestras de cada conjunto en las c clases, calculándose $\hat{\Delta}^{SE}(c)$ y $\hat{\Delta}^{VC}(c)$ según la ecuación (5.5) para cada partición generada con esas c clases. A partir de la observación de los valores obtenidos, se deduce que cuando se alcanza el verdadero número de clases según la tabla 5.1, se verifica que tanto $\hat{\Delta}^{SE}(c)$ como $\hat{\Delta}^{VC}(c)$ son siempre menores que 0.1. Se toma, por tanto, este valor como umbral para estimar el mejor número de clases para un conjunto de muestras cuyo número de clases resulta desconocido. De este modo se llega a los procedimientos no supervisados descritos en el capítulo cuatro. Por consiguiente, dado un conjunto de muestras de las que no se conoce su número de clases, la mejor partición se establece con un número de clases tal que se cumple la relación dada por la ecuación siguiente,

$$\hat{\Delta}^f(c) \leq U \quad (5.6)$$

donde U es el umbral establecido según lo anterior, fijado al valor de 0.1.

Como el comportamiento de $\hat{\Delta}^{SE}(c)$ y $\hat{\Delta}^{VC}(c)$ son similares, para la estimación del número de clases se usará el primero para el conjunto de imágenes aéreas utilizadas en nuestros experimentos. Su elección viene determinada por su ligero mejor comportamiento, según se desprende del estudio realizado en la sección 3.4.4.

5.3.2 Fase de decisión

5.3.2.1 Mapas Cognitivos Fuzzy y relajación

El proceso de actualización de los estados de los nodos en los *MCFs* es iterativo. Como se describe en la sección 4.4.3.6, el proceso requiere fijar los parámetros ε y t_{max} que representan respectivamente el nivel de variación de los valores de los nodos entre dos iteraciones consecutivas y el máximo número de iteraciones a considerar.

Tal y como se ha especificado en la descripción de los *MCFs*, éste se aplica sólo a datos que cumplan los principios de proximidad, similitud y conectividad de la Gestalt, lo cual ocurre sólo en los datos de la tabla 5.1, que son exactamente imágenes. Pues bien, para determinar los valores de los parámetros anteriores, se toman como referencia las imágenes correspondientes a las categorías de datos 13, 14, 15 y 16 de dicha tabla; a partir de los cuales se determina que con $\varepsilon = 0.02$ y $t_{max} = 20$ los resultados obtenidos son satisfactorios a juicio del experto humano.

El razonamiento anterior es igualmente válido para el caso del método global de relajación *REL*, descrito en la sección 4.5.

5.3.2.2 Operador de agregación de Yager

Según la ecuación (3.57) el operador de agregación de Yager depende del parámetro p , que debe ser igualmente ajustado. Con tal propósito y una vez estimados los parámetros de los clasificadores individuales, se utilizan los dieciséis conjuntos de validación V_i de la tabla 5.1, de forma que el mejor valor de p se elige en función del menor error obtenido en la clasificación de los datos de dichos conjuntos variando p desde 1 hasta 10

con incrementos de 0.2. De esta forma se determina que el mejor valor para dicho parámetro es 4.0.

5.3.3 Resumen de parámetros

Como colofón a los procesos de ajuste de parámetros descritos previamente, en la tabla 5.3 se resumen los valores estimados de los mismos para cada uno de los métodos propuestos. Se distingue entre fase de entrenamiento y fase de decisión según que los parámetros estén involucrados en una u otra fase. En realidad, en la fase de entrenamiento, sólo se requiere el ajuste de parámetros para los métodos de clasificación individuales, siendo en la fase de decisión donde se determinan los correspondientes para los métodos combinados *MCFs* y *REL*.

Tabla 5.3 Resumen de parámetros

Fase	Clasificadores	Parámetros y sus valores
Entrenamiento	Clasificador borroso (<i>FC</i>)	$m = 2.1$; $\varepsilon = 0.01$ y $t_{max} = 50$
	Ventana de Parzen (<i>BN</i>)	$r = 0.3$; $h = 1.1383$
	Mapas Auto-Organizativos (<i>SO</i>)	$\alpha_{final} = \alpha_{inicial} = 4.38$, $\varepsilon = 0.01$ y $t_{max} = 50$
	Cuantización vectorial (<i>L1, L2</i>)	$\varepsilon = 0.01$; $t_{max} = 50$; $\sigma = 0.08$
Decisión	Mapas Cognitivos Fuzzy (<i>MCFs</i>)	$\varepsilon = 0.02$; $t_{max} = 20$
	Operador de agregación de Yager	$p = 4.0$

5.4 Diseño de una estrategia general de pruebas

Como se ha mencionado reiteradamente, todos los clasificadores descritos en este trabajo, incluyendo las estrategias combinadas, incluyen tanto una fase de *entrenamiento* como una de *decisión*. Los métodos de clasificación individuales se entrenan para después combinarlos convenientemente durante la fase de decisión. Parece lógico pensar, por tanto, que el entrenamiento debe influir necesariamente en el proceso. En consecuencia, para comprobar cómo afecta el aprendizaje a los resultados, se ha diseñado la estrategia general de pruebas, que se expone a continuación, basada en cuatro pasos. En cada paso se realiza un aprendizaje incremental con respecto al paso previo, cuyo objetivo consiste en verificar, cómo a medida que aumenta el aprendizaje se obtienen mejores resultados en la clasificación de las nuevas muestras. Esto tiene su

explicación lógica, ya que cuanto mayor es el conocimiento adquirido por el sistema, mejor y más fácil resulta la decisión a tomar.

Un segundo objetivo de la estrategia de pruebas diseñada consiste en verificar el mejor comportamiento de los métodos combinados frente a las estrategias individuales y dentro de los combinados, comprobar el comportamiento de las estrategias propuestas en este trabajo de investigación frente al resto de las combinaciones.

Los cuatro pasos mencionados son los que se describen a continuación, en ellos se procesan los conjuntos de imágenes SI (inicial), S0, S1, S2 y S3 descritos en la sección 5.2.2 como se indica seguidamente. Por consiguiente, en esta fase se utilizan las imágenes aéreas descritas en la mencionada sección.

a) PASO 0 (inicial):

El conjunto de imágenes procesadas en este paso es exactamente el SI con las 12 imágenes seleccionadas aleatoriamente del conjunto de las 52 disponibles. El objetivo de este paso consiste básicamente en estimar el mejor número de clases existente en dichas imágenes a partir de las muestras disponibles, procedentes de ellas. Para obtener las muestras se realiza un submuestreo por dos sobre cada imagen, que consiste en tomar una fila y una columna de cada dos. Por consiguiente, dado que las imágenes originales son de 512×512 píxeles y que las muestras son píxeles con sus componentes espectrales como propiedades, de cada imagen se obtienen 256×256 muestras, resultando en total $12 \times 256 \times 256 = 786432$ muestras para este primer paso inicial.

Con estas muestras se inicia el proceso de partición de las mismas en c clases, para ello se aplica el procedimiento no supervisado descrito en la sección 4.2.1 o equivalentemente en la sección 4.4.1. El procedimiento es que se describe en las mencionadas secciones, sintetizado como sigue:

1. Desde $c = 1$ hasta $c = 8$ realizar el proceso de partición de las muestras en clases mediante el procedimiento pseudo-aleatorio descrito en la sección 3.3.1.1.
2. Obtener las variaciones relativas de SE para cada partición obtenida con el actual c , según las ecuaciones (5.5) o equivalentemente su precursora (3.43).

3. Determinar si la partición es válida o no en función de que se verifique o no la relación de la ecuación (5.6).
4. Si la partición es válida, el número de clases para el conjunto de muestras dado se establece el valor de c correspondiente.
5. Conocido el número de clases, las muestras quedan definitivamente asignadas a las respectivas clases, de este modo se pueden estimar los parámetros de aprendizaje asociados con los diferentes clasificadores individuales mediante los correspondientes procesos de *entrenamiento*, descritos en el capítulo tres. Los parámetros aprendidos se almacenan en la BC .

b) PASO 1

A diferencia del paso previo, donde sólo se realiza la fase de entrenamiento, en éste se lleva a cabo en primer lugar un proceso de decisión para clasificar los píxeles de las imágenes de los conjuntos S_0 y S_1 . En segundo lugar se realiza un nuevo proceso de entrenamiento como se describe seguidamente.

Dados los conjuntos S_0 y S_1 , los píxeles pertenecientes a dichas imágenes se clasifican como pertenecientes a alguna de las c clases obtenidas en el PASO 0. Tras lo cual se calculan los errores de clasificación por comparación de los resultados obtenidos con las correspondientes imágenes de referencia, estas últimas se obtienen según se describe posteriormente en la sección 5.6.

Las muestras clasificadas procedentes del conjunto S_1 se añaden como nuevas muestras de entrenamiento a las ya existentes provenientes del conjunto S_1 . Con todas estas nuevas muestras de entrenamiento se procede a realizar un nuevo proceso de entrenamiento con todos los clasificadores individuales, naturalmente manteniendo el mismo número de clases \hat{c} ya prefijado. En este nuevo proceso de entrenamiento se estiman de nuevo los parámetros de aprendizaje involucrados en los clasificadores individuales, que quedan finalmente almacenados en la BC y ésta actualizada consecuentemente.

Las muestras clasificadas procedentes de S_0 no participan en este nuevo proceso de entrenamiento, esto se ha decidido así intencionadamente con el fin de verificar a través del conjunto testigo S_0 cómo influye el proceso de aprendizaje sobre él durante su

clasificación en los pasos posteriores. De esta forma, al no utilizarse dicho conjunto durante el entrenamiento, se evita su propia influencia en su clasificación.

c) PASO 2:

En este paso se procede del mismo modo que en el PASO 1, si bien utilizando los conjuntos S0 y S2. Ahora los resultados de la clasificación procedentes de S2 se añaden a las muestras de entrenamiento disponibles con el fin de realizar un nuevo proceso de entrenamiento; tras lo cual, los nuevos parámetros aprendidos quedan almacenados en la BC. Como en el paso previo, las muestras clasificadas procedentes de S0 no se utilizan en el entrenamiento.

d) PASO 3:

En este paso se procede del mismo modo que en los PASOS 1 y 2, si bien utilizando los conjuntos S0 y S3, de forma que de nuevo se clasifican las muestras procedentes de estos conjuntos, no obstante, ya no se realizan más procesos de entrenamiento.

Como se ha mencionado previamente, en los PASOS 1 y 2 se añaden nuevas muestras para realizar un nuevo proceso de entrenamiento. El número de muestras añadidas en cada uno de estos dos pasos resulta ser $10 \times 512 \times 512 = 2621440$, que se corresponde con el resultado de multiplicar el número de píxeles de cada imagen por el número de imágenes en cada conjunto S1 y S2. El número total de muestras de entrenamiento utilizadas durante el proceso de pruebas es, por tanto, la suma de muestras procedentes de las imágenes del conjunto SI, más las procedentes de los conjuntos S1 y S2 que se introducen en los PASOS 1 y 2 respectivamente, haciendo un total de $12 \times 256 \times 256 + 2 \times 10 \times 512 \times 512 = 6029312$ muestras.

El proceso de pruebas descrito previamente se sintetiza en la figura 5.4, según las dos fases contempladas: *decisión* y *entrenamiento*. En él se sigue la numeración indicada alternando ambas fases de entrenamiento y decisión según se indica:

- 1) *Entrenamiento* con las muestras procedentes de SI, se estima el número de clases más apropiado c y la partición de los datos en esas c clases.
- 2) *Clasificación* de las muestras proporcionadas por los conjuntos S0 y S1.

- 3) *Entrenamiento* con las muestras procedentes de SI y S1
- 4) *Clasificación* de las muestras proporcionadas por los conjuntos S0 y S2.
- 5) *Entrenamiento* con las muestras procedentes de SI, S1 y S2
- 6) *Clasificación* de las muestras proporcionadas por los conjuntos S0 y S3.

Tras cada proceso de clasificación se procede a determinar los porcentajes de error obtenidos en ese momento para cada imagen por comparación con las imágenes de referencia. Por tanto, como puede observarse e indicado previamente, los errores del conjunto S0 se computan tres veces; esto es, en los PASOS 1, 2 y 3, con el fin de determinar su evolución según progresa el aprendizaje.

Por otro lado, tras cada proceso de entrenamiento, los parámetros estimados por cada uno de los clasificadores se almacenan en la *BC* para su posterior recuperación durante la siguiente fase de decisión.

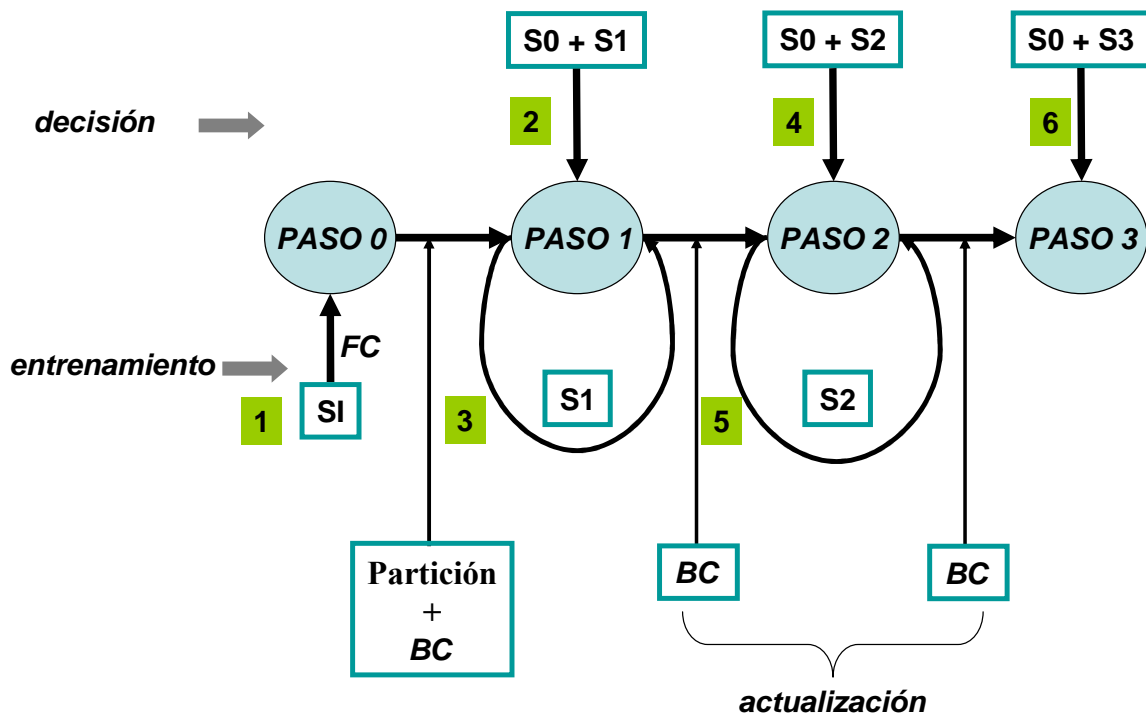


Figura 5.4 Síntesis de la estrategia de prueba

En la sección 5.5 se describe la fase de entrenamiento inicial realizada en el PASO 0, que incluye la estimación del mejor número de clases. En la sección 5.6 se describe el proceso para generar las imágenes de referencia. Posteriormente, en la sección 5.7 se

analizan los resultados obtenidos en los PASOS 1 a 3 por comparación con esas imágenes de referencia.

5.5 Entrenamiento inicial

Llegados a este punto, el objetivo que se plantea es doble, por un lado se trata de estimar el mejor número de clases subyacente en el conjunto de imágenes SI, lo que permitirá a su vez obtener las competencias de los seis clasificadores individuales involucrados: *FC*, *BP*, *BN*, *SO*, *L1* y *L2*. Por otro lado, se trata de estimar los parámetros de aprendizaje según esos mismos seis clasificadores y reflejados en la tabla 3.1. En el primer caso se sigue el procedimiento sintetizado en la sección previa en la descripción del PASO 0; mientras que para el segundo, se aplican los procesos de aprendizaje descritos en la sección 3.3, para cada uno de los seis clasificadores durante las fases de entrenamiento.

5.5.1 Estimación inicial del número de clases y competencias de los clasificadores

Siguiendo el proceso descrito en la sección 4.2.3 cada una de las 12 imágenes del conjunto SI constituye uno de los B subconjuntos $\{S_1, S_2, \dots, S_B\}$, siendo $B = 12$.

Seguidamente, para cada subconjunto S_i , se calcula $\hat{\Delta}^{SE}(c)$, de acuerdo con la ecuación (5.5), variando c desde 1 hasta $G-1$, con G igual a 8. Cuando se verifica la relación dada en la ecuación (5.6), coincidiendo con un número de clases tal que $\hat{c} = c$, este número se considera el mejor para ese conjunto que llamamos \hat{c}_i para asociarlo con S_i . Antes de continuar aclaremos algunas cuestiones relacionadas mediante un sencillo ejemplo, la tabla 5.4 muestra el comportamiento de $\hat{\Delta}^{SE}(c)$ y $\hat{\Delta}^{VC}(c)$ para el conjunto S_1 con las particiones obtenidas mediante el clasificador *FC* y por tanto, con los centros para *SE*. En la figura 5.5 se muestran gráficamente los mismos resultados, observándose cómo tanto $\hat{\Delta}^{SE}(c)$ como $\hat{\Delta}^{VC}(c)$ disminuyen rápidamente hasta alcanzar $\hat{c}_1 = 4$, a partir del cual, la disminución es mucho más suave. Por tanto, aplicando la regla (5.6) para este conjunto y teniendo en cuenta el valor del umbral $U = 0.1$, obtenido por el procedimiento descrito en la sección 5.3.1.5, se determina que la mejor estimación para el número de clases es $\hat{c}_1 = 4$.

Obsérvese en la figura 5.5 un comportamiento similar tanto para $\hat{\Delta}^{SE}(c)$ como para $\hat{\Delta}^{VC}(c)$, verificándose en ambos casos la coincidencia en la estimación del número de clases, en este caso para $\hat{c}_i = 4$. Este similar comportamiento es aplicable al resto de los subconjuntos S_i , motivo por el cual se justifica la utilización sólo de $\hat{\Delta}^{SE}(c)$ en la evaluación de la ecuación (5.6).

Tabla 5.4 Comportamiento de los coeficientes $\hat{\Delta}^{SE}(c)$ y $\hat{\Delta}^{VC}(c)$ para S_1

número de clases (c)	1	2	3	4	5	6	7
Δ	c_1-c_2	c_2-c_3	c_3-c_4	c_4-c_5	c_5-c_6	c_6-c_7	c_7-c_8
$\hat{\Delta}^{SE}(c)$	0.4478	0.2408	0.1203	0.0661	0.0600	0.0338	0.0311
$\hat{\Delta}^{VC}(c)$	0.4842	0.2242	0.1149	0.0632	0.0546	0.0316	0.0273

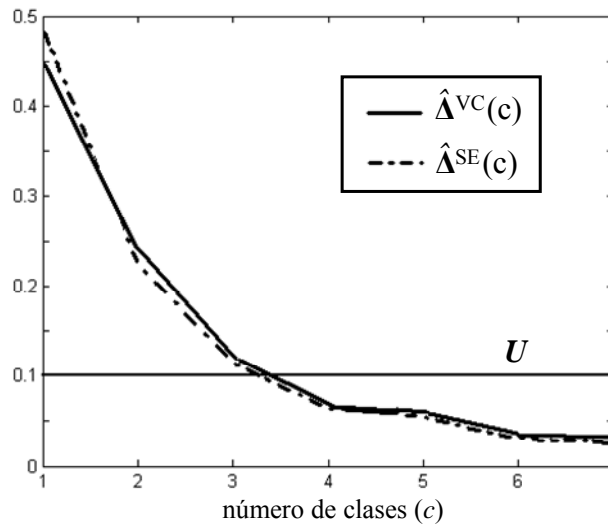


Figura 5.5 Valores de $\hat{\Delta}^{SE}(c)$ y $\hat{\Delta}^{VC}(c)$ frente al número de clases

En la tabla 5.5 se muestra el número de clases óptimo para cada uno de los doce subconjuntos disponibles S_i , según el procedimiento descrito.

Tabla 5.5 Estimación del número de clases para cada subconjunto S_i

subconjuntos	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
\hat{c}_i	4	3	4	3	4	4	5	4	4	3	4	4

A partir de los resultados mostrados en dicha tabla, se observa que el valor cuatro es el más frecuente, por tanto, utilizando un criterio de votación mayoritaria, este es el valor finalmente elegido como el mejor número global de clases estimado para el conjunto inicial de imágenes SI, siendo a la vez el número que se utiliza a partir de este momento

para llevar a cabo todos los procesos de entrenamiento y decisión; en consecuencia, de aquí en adelante el número estimado de clases para las imágenes disponibles es $\hat{c} = 4$.

En el método combinado propuesto en el capítulo cuatro, relativo al proceso de decisión basado en el paradigma *MCDMF*, durante la clasificación se tienen en cuenta las competencias absolutas otorgadas a cada clasificador individual, que interviene en la combinación. Esto se realiza mediante el proceso descrito en la sección 4.2.3 a través de las ecuaciones (4.12) y (4.13), que se reproducen aquí bajo las ecuaciones (5.7) y (5.8) respectivamente.

$$\alpha_{l_1}^{SE}(\hat{c}_i) = 1 - \hat{\Delta}^{SE}(\hat{c}_i) \text{ y } \alpha_{l_2}^{VC}(\hat{c}_i) = 1 - \hat{\Delta}^{VC}(\hat{c}_i) \quad (5.7)$$

$$\alpha_m = \left| \sum_{i=1}^B (-1)^s \alpha_m^f(\hat{c}_i) \right| \quad (5.8)$$

El subíndice l_1 se refiere a los clasificadores *FC*, *BP*, *SO*, *LI*, *L2* y l_2 exclusivamente a *BN*; f es *VC* para *BN* y *SE* para los clasificadores restantes; con $s = 0$ si $\hat{c}_i = \hat{c}$ y $s = 1$ si $\hat{c}_i \neq \hat{c}$; m se refiere a todos los clasificadores incluidos tanto en l_1 como en l_2 . Como se mencionó en la sección 4.2.3, para un clasificador dado l_1 , $\hat{\Delta}^{SE}(\hat{c}_i)$ se calcula teniendo en cuenta los centros proporcionados por dicho clasificador l_1 para la partición con \hat{c}_i clases y la partición con $\hat{c}_i + 1$ clases. Además, $\hat{\Delta}^{VC}(\hat{c}_i)$ se calcula a partir de las mismas particiones anteriores teniendo en cuenta la dispersión de las muestras en ellas.

En la tabla 5.6 se muestran los valores $\alpha_{l_1}^{SE}(\hat{c}_i)$ y $\alpha_{l_2}^{VC}(\hat{c}_i)$ para los diferentes clasificadores de acuerdo con cada uno de los conjuntos S_i , para los cuales se vuelve a mostrar el número de clases estimadas según lo expresado previamente y reflejadas en la tabla 5.5. En la columna α_m se muestran los resultados obtenidos por la aplicación de la expresión (5.8) para cada uno de los valores individuales obtenidos para los correspondientes conjuntos S_i . Siguiendo los razonamientos expresados en la sección 4.2.4, a través de la ecuación (5.9), se obtienen las competencias absolutas promediadas α^1 y α^2 asociadas respectivamente con los criterios CR_1 y CR_2 ,

$$\alpha^1 = (\alpha_{FC} + \alpha_{BP} + \alpha_{BN})/3; \quad \alpha^2 = (\alpha_{SO} + \alpha_{LI} + \alpha_{L2})/3 \quad (5.9)$$

Finalmente, se calculan los pesos específicos (z_1, z_2) asociados a cada criterio,

$$z_1 = \alpha^1 / (\alpha^1 + \alpha^2); \quad z_2 = \alpha^2 / (\alpha^1 + \alpha^2) \quad \text{con} \quad z_1 + z_2 = 1 \quad (5.10)$$

A la vista de los resultados mostrados en la tabla 5.6, se deduce que los pesos específicos para ambos criterios son muy similares, lo que se traduce en el hecho de que los dos criterios son valorados por igual durante la fase de decisión; esto, incluso a pesar de que a nivel individual las diferencias se hacen más significativas entre algunos clasificadores, tal es el caso observado entre los clasificadores *BN* y *BP* e incluso *BN* y *SO*.

En la tabla 5.6 se muestran también las densidades fuzzy $g^m = \alpha_m / \sum_{k \in \Omega} \alpha_k$ para cada clasificador $\Omega = \{FC, BP, BN, SO, L1, L2\}$, que son utilizadas por el método combinado de clasificación *IF* descrito en la sección 4.3.

Tabla 5.6 Variación de las funciones criterio $\alpha_{FC}^{SE}(\hat{c}_i)$ para cada clasificador m y para cada subconjunto S_i ; competencias absolutas promediadas de cada clasificador α_m ; competencias medias α^1 y α^2 para cada criterio CR_1 y CR_2 ; pesos específicos de cada criterio z_1 y z_2 ; densidades fuzzy g^m para cada clasificador

			S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}				
		\hat{c}_i	4	3	4	3	4	4	5	4	4	3	4	4	α_m	g^m		
Clasificadores individuales	FC	$\alpha_{FC}^{SE}(\hat{c}_i)$	0.92	0.91	0.93	0.92	0.95	0.93	0.91	0.95	0.91	0.93	0.94	0.97	3.83	0.17	α^1	$CR_1(z_1)$
	BP	$\alpha_{BP}^{SE}(\hat{c}_i)$	0.95	0.91	0.92	0.94	0.96	0.90	0.93	0.97	0.96	0.92	0.94	0.94	3.84	0.17		
	BN	$\alpha_{BN}^{VC}(\hat{c}_i)$	0.91	0.93	0.93	0.97	0.92	0.91	0.94	0.91	0.90	0.92	0.91	0.92	3.55	0.16		
	SO	$\alpha_{SO}^{SE}(\hat{c}_i)$	0.92	0.95	0.96	0.92	0.93	0.96	0.96	0.99	0.95	0.94	0.92	0.95	3.81	0.17	α^2	$CR_2(z_2)$
	$L1$	$\alpha_{L1}^{SE}(\hat{c}_i)$	0.91	0.96	0.94	0.93	0.92	0.95	0.97	0.98	0.93	0.93	0.91	0.93	3.68	0.16		
	$L2$	$\alpha_{L2}^{SE}(\hat{c}_i)$	0.92	0.94	0.91	0.92	0.93	0.92	0.93	0.97	0.91	0.92	0.92	0.92	3.69	0.16		

5.5.2 Aprendizaje de los parámetros de cada clasificador

De acuerdo con lo explicado en el capítulo tres, los parámetros estimados o aprendidos son, por un lado, los centros de las clases o valores medios identificándose como sigue, donde se ha obviado el subíndice j que referencia la respectiva clase w_j : ν para FC según la ecuación (3.4), m para BP de acuerdo con la ecuación (3.18) y c para SO , $L1$ y $L2$ de acuerdo con las ecuaciones (3.33), (3.35) y (3.37) respectivamente. Por otro lado, en el caso de los clasificadores BP y BN también se aprenden las matrices de covarianza asociadas con cada una de las clases, obtenidas según la ecuación (3.18). Como ya sabemos, el clasificador BN no aprende más parámetros que las mencionadas matrices de covarianza, ya que estima las funciones de densidad de probabilidad con todas las muestras de clase disponibles según la ecuación (3.23).

En la tabla 5.7 aparece el número de muestras distribuidas en las cuatro clases para el total de los doce subconjuntos S_i , pertenecientes al conjunto inicial SI así como los valores de los centros según los diferentes clasificadores. Estos valores se han obtenido para el total de las 756432 muestras disponibles procedentes de SI y una vez que se ha determinado la existencia de las cuatro clases. En la tabla 5.8 se muestran los valores de las matrices de covarianza obtenidas mediante el estimador no paramétrico de máxima verosimilitud según la ecuación (3.18).

Tabla 5.7 Número de muestras y centros de las clases para los clasificadores individuales FC , BN , SO , $L1$ y $L2$ obtenidos a partir del conjunto de datos SI

	clase 1 (w_1)	clase 2 (w_2)	clase 3 (w_3)	clase 4 (w_4)
Número de muestras	51713	388360	186570	129790
FC (ν)	(229.1, 194.0, 184.4)	(93.0, 106.4, 6.5)	(168.0, 142.8, 108.6)	(35.3, 28.8, 19.9)
BN (m)	(226.7, 191.9, 180.4)	(93.1, 106.0, 66.4)	(167.0, 142.6, 108.4)	(37.5, 31.3, 21.5)
SO (c)	(228.1, 195.3, 186.9)	(85.6, 110.2, 67.4)	(171.3, 144.3, 110.1)	(38.9, 33.8, 19.3)
$L1$ (c)	(230.2, 198.3, 189.3)	(88.7, 111.3, 68.6)	(182.1, 158.7, 124.1)	(39.3, 35.9, 21.1)
$L2$ (c)	(233.4, 201.3, 199.3)	(89.8, 112.1, 88.7)	(183.3, 155.6, 121.3)	(40.1, 36.2, 21.9)

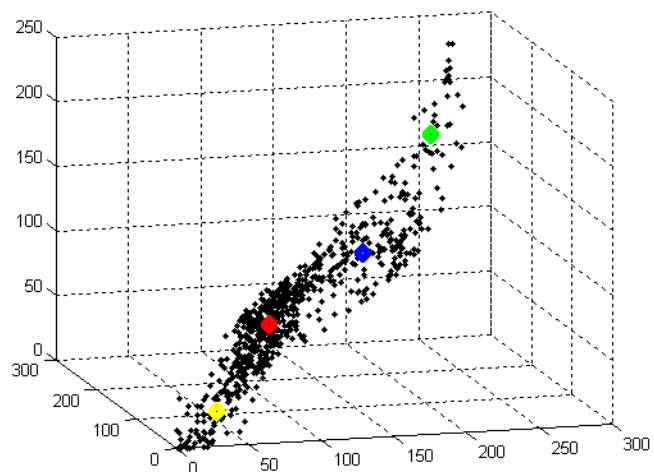
Tabla 5.8 Matrices de covarianza para el estimador paramétrico BP según las clases

<i>clase 1 (w_1)</i>	<i>clase 2 (w_2)</i>	<i>clase 3 (w_3)</i>	<i>clase 4 (w_4)</i>
$\begin{bmatrix} 342.1 & 254.8 & 457.8 \\ 254.8 & 467.4 & 548.9 \\ 457.8 & 548.9 & 889.5 \end{bmatrix}$	$\begin{bmatrix} 279.8 & 197.5 & 178.6 \\ 197.5 & 384.3 & 211.5 \\ 178.6 & 211.5 & 167.0 \end{bmatrix}$	$\begin{bmatrix} 603.9 & 97.7 & 212.9 \\ 97.7 & 208.7 & 172.3 \\ 212.9 & 172.3 & 217.8 \end{bmatrix}$	$\begin{bmatrix} 396.0 & 346.6 & 257.1 \\ 346.6 & 518.9 & 249.9 \\ 257.1 & 249.9 & 214.5 \end{bmatrix}$

En la figura 5.6(a) se muestra la imagen correspondiente al subconjunto S_1 , mientras que en (b) se visualiza una serie de muestras representativas de las cuatro clases obtenidas con los 12 conjuntos S_i . Sólo se representan 746 muestras para evitar problemas derivados del solapamiento en la representación gráfica. Esta representación tridimensional es la que se corresponde con las tres componentes espectrales R,G, B de los píxeles, los cuales, como ya se ha mencionado reiteradamente, constituyen las muestras. En dicha figura se identifican los centros asociados con cada clase según se indica por los correspondientes marcadores. Los centros mostrados son los que corresponden al clasificador BP. Para una mejor visualización, en la figura 5.7 se muestra la misma distribución de datos proyectados al espacio bidimensional en lugar del tridimensional. Esta proyección se ha realizado aplicando la técnica de componentes principales, que considera las dos componentes más relevantes en función de los autovalores más significativos de los datos (Pajares y Cruz, 2007).



(a)



(b)

Figura 5.6 (a) imagen de textura natural correspondiente al conjunto S_1 ; (b) distribución tri-dimensional de las muestras con sus centros de clase correspondientes

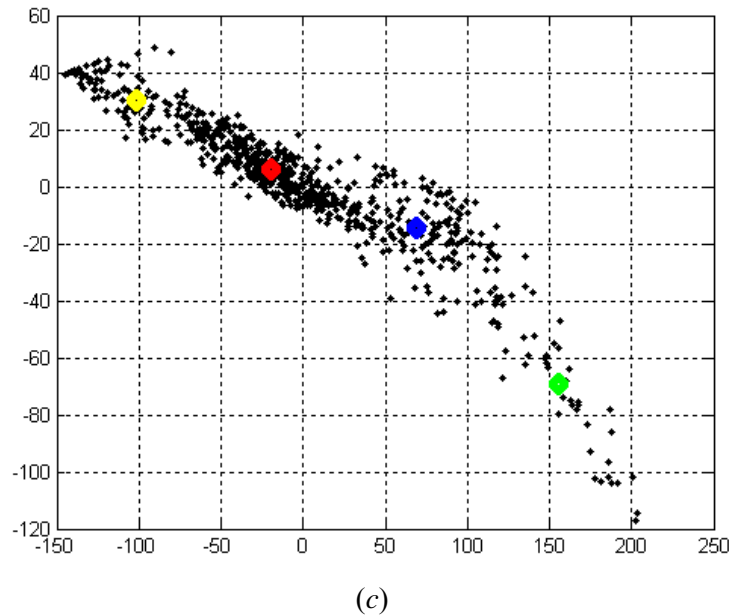


Figura 5.7 Distribución bi-dimensional de los mismos datos representados en la figura 5.6 (b)

5.6 Generación de las clases de referencia

Con el fin de verificar el comportamiento de los diferentes métodos, tanto individuales como combinados, es necesario construir la base de imágenes de referencia con las que se han de comparar los resultados. Técnicamente, en la literatura especializada, estas imágenes de referencia se conocen como “ground truth”. El procedimiento para elaborar cada imagen de referencia se describe a continuación:

- 1) Se parte del conocimiento del número de clases, que se ha establecido según el proceso descrito en la sección 5.5.1, resultando ser cuatro en nuestro caso.
- 2) Se procede a clasificar la imagen sobre la que se van a establecer las clases de referencia mediante el clasificador *BP*, por ser el clasificador individual que mejores resultados proporciona, según se verá posteriormente.
- 3) Se crea una imagen binaria para cada clase, de suerte que en cada una de dichas imágenes binarias, los píxeles de la correspondiente clase se etiquetan como píxeles en blanco (con valores de intensidad de 255) mientras que los píxeles pertenecientes al resto de clases se etiquetan en negro (con valores de intensidad cero).
- 4) Las imágenes binarias se retocan manualmente siguiendo las indicaciones de los expertos destinatarios de las mismas hasta que éstos consiguen un alto grado de

satisfacción. En nuestro caso, los expertos son los técnicos responsables de la empresa DIMAP (2008), que da origen a la motivación principal de esta investigación.

- 5) Las imágenes binarias, así obtenidas, se almacenan convenientemente, de forma que durante los correspondientes procesos de decisión se utilicen como testigos con los que determinar el porcentaje de errores de clasificación.

A continuación se ilustra el proceso descrito según los cinco puntos anteriores. En la figura 5.8(a) se muestra una de las imágenes originales perteneciente al conjunto S0; en (b) aparece la correspondencia entre los colores que identifican las clases sobre la imagen original y los colores sintéticos asignados para etiquetar cada clase. Los colores que representan las clases, según sus texturas originales, corresponden a los valores espectrales en el modelo de color RGB para los centros de cada clase obtenidos mediante el clasificador BP. En la figura 5.8(c) se muestra la imagen etiquetada de acuerdo con la correspondencia del mapa de colores definido en (b). La correspondencia entre dichas etiquetas y las texturas naturales es la siguiente:

- *Amarillo* con masas de vegetación boscosa
- *Azul* con campos de cultivo en barbecho
- *Verde* con vegetación de pradera y campos de cultivo sembrados
- *Rojo* con construcciones artificiales y caminos de tierra

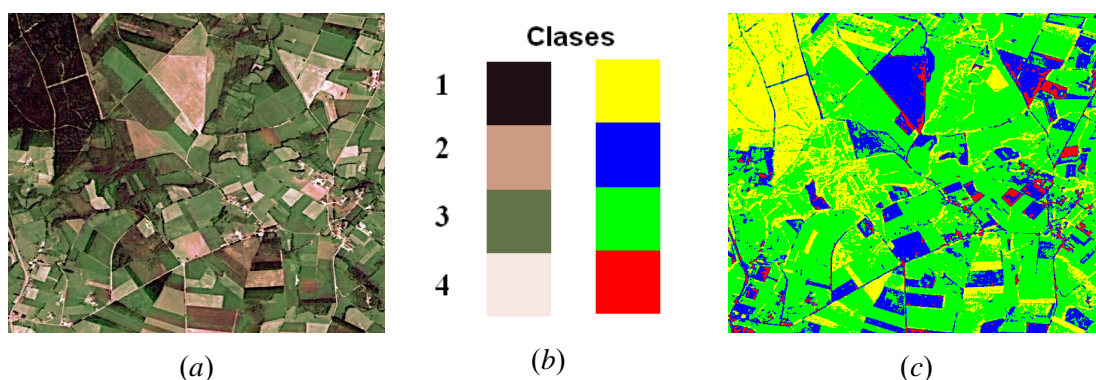


Figura 5.8 (a) imagen original perteneciente al conjunto S0; (b) correspondencia entre etiquetas y clases; (c) imagen etiquetada en sus cuatro clases de acuerdo a la equivalencia establecida en (b)

En las figuras 5.9 a 5.12 se muestran las imágenes de referencia para las cuatro clases establecidas. Las imágenes de la serie (a) representan la binarización obtenida directamente a partir de los resultados proporcionados por el clasificador BP. Así por

ejemplo, para el caso de la figura 5.9(a) los píxeles en blanco proceden de la transformación de los píxeles de color amarillo, que aparecen en la imagen de la figura 5.8(c) a blancos, dejando el resto a negros. Las imágenes de la serie (b) son las obtenidas mediante el retoque guiado por los expertos a partir de los correspondientes resultados de la serie (a). Por tanto, la serie (b) constituye el conjunto de imágenes de referencia o ground-truth para la imagen original dada en la figura 5.8(a).

Este proceso se repite para todas y cada una de las imágenes pertenecientes a los conjuntos S0, S1, S2 y S3, sometidos al proceso de clasificación, obteniendo por tanto, un total de 160 imágenes de referencia como resultado del producto de cuatro conjuntos por diez imágenes en cada conjunto y con cuatro clases por imagen.

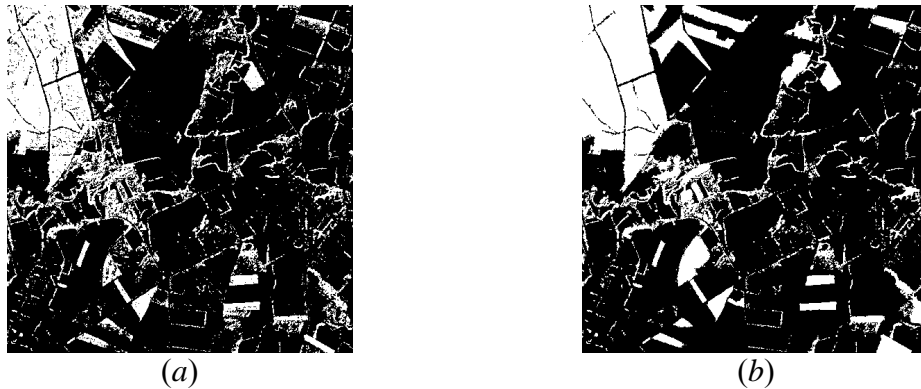


Figura 5.9 Clase 1: (a) imagen binaria obtenida mediante el clasificador BP; (b) imagen de referencia

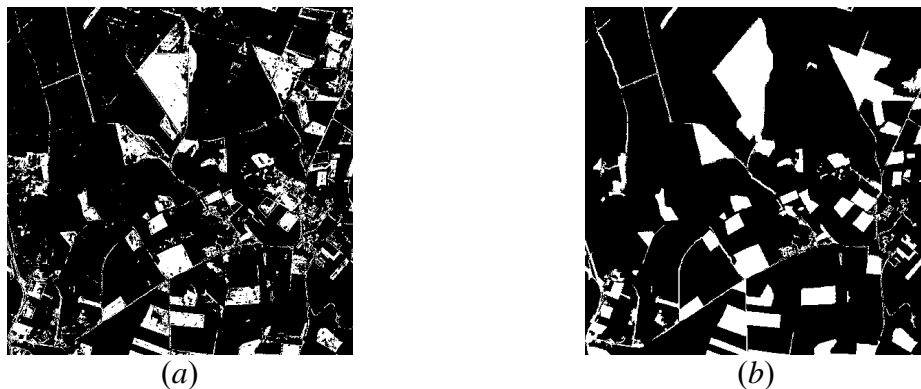


Figura 5.10 Clase 2: (a) imagen binaria obtenida mediante el clasificador BP; (b) imagen de referencia

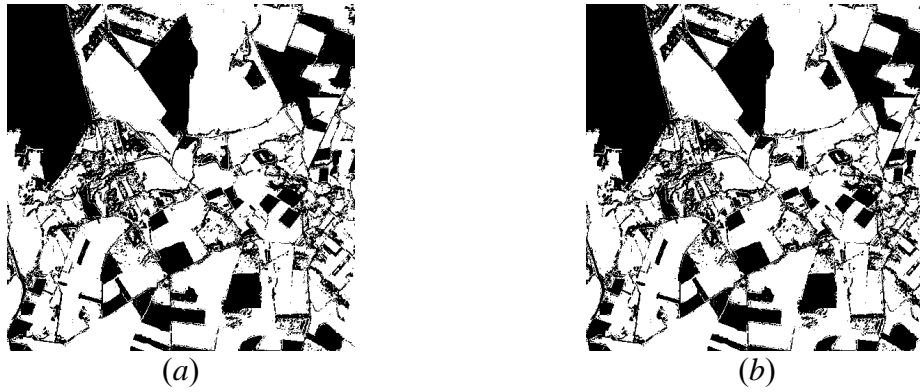


Figura 5.11 Clase 3: (a) imagen binaria obtenida mediante el clasificador BP; (b) imagen de referencia

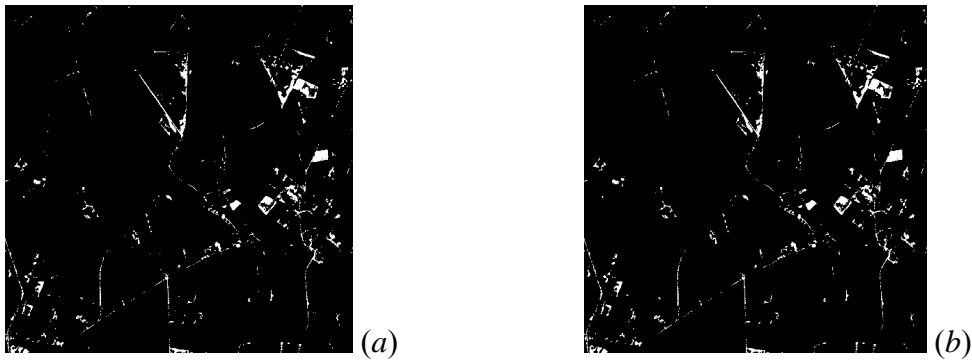


Figura 5.12 Clase 4: (a) imagen binaria obtenida mediante el clasificador BP; (b) imagen de referencia

5.7 Resultados obtenidos

Según la estrategia de pruebas diseñada de acuerdo con los cuatro PASOS descritos en la sección 5.4, el objetivo ahora se centra en evaluar el comportamiento de los diferentes métodos de clasificación tanto individuales como combinados y dentro de éstos los propuestos en el capítulo cuatro, que constituyen la esencia de la investigación. En primer lugar, para poder llevar a cabo la evaluación es necesario calcular el porcentaje de error en la clasificación de las imágenes contenidas en los conjuntos S0, S1, S2 y S3, lo que se realiza en la sección 5.7.1. En segundo lugar, se analizan dichos resultados, según su comportamiento general, esta tarea se aborda en la sección 5.7.2. En tercer lugar, se analizan los tiempos de cómputo requeridos por los diferentes métodos, haciendo especial hincapié en aquellos de naturaleza iterativa, sección 5.7.3. En cuarto lugar, en la sección 5.7.4 se realiza un análisis comparativo de resultados, comenzando por la influencia debida al aprendizaje adquirido por el sistema durante su entrenamiento y continuando por el estudio de los diferentes métodos, tanto individuales como combinados en todas sus vertientes. Finalmente, en quinto y último

lugar, en la sección 5.7.5 se muestran algunos resultados relevantes dignos de consideración.

5.7.1 Cómputo del porcentaje de error

En cada uno de los cuatro PASOS se procesan dos conjuntos de imágenes para los que se obtienen los errores de clasificación correspondientes. Así en el PASO 1 se computan los errores por separado para los conjuntos S0 y S1; en el PASO 2 para S0 y S2 y finalmente en el PASO 3 para los conjuntos S0 y S3, en todos los casos siguiendo la estrategia de pruebas definida previamente en la sección 5.4. De este modo, se identifica cada conjunto como SN donde $N = 0,1,2,3$, refiriéndonos así a los respectivos conjuntos S0, S1, S2 y S3 según el valor N .

Por otro lado, cada conjunto SN está formado por diez imágenes de texturas naturales. Esto nos permite identificar cada imagen en cada subconjunto como I_N^r , siendo N el indicativo del subconjunto al que pertenece la imagen y r su número de orden dentro del subconjunto, $r = 1, \dots, R$ y por tanto con $R = 10$. Por otro lado, i se refiere al píxel dado en la localización espacial (x,y) para una imagen I_N^r . Se define un contador de error E_N^r para cada imagen r en el conjunto SN , inicializándose a cero en cada PASO.

Teniendo en cuenta el correspondiente proceso de decisión para cada clasificador, según se ha descrito en los capítulos tres y cuatro, el píxel i será clasificado como perteneciente a una de las clases w_j , en cuyo caso, $i \in w_j$. Si ese mismo píxel en la correspondiente imagen binaria de referencia (ground-truth) asociada a la clase w_j , es un píxel etiquetado como negro, entonces se considera que éste ha sido incorrectamente clasificado, en cuyo caso el contador de error se incrementa en una unidad, es decir $E_N^r = E_N^r + 1$. Siempre se mantiene un contador por cada clasificador individual o combinado. La razón de error de la imagen I_N^r , con un número de píxeles Z , resulta ser: $e_N^r = E_N^r / Z$; en nuestro caso $Z = 512 \times 512$. La razón media de error para el conjunto SN en cada PASO, así como la desviación estándar, vienen dadas por \bar{e}_N y $\bar{\sigma}_N$ respectivamente según la ecuación siguiente,

$$\bar{e}_N = \frac{1}{R} \sum_{r=1}^R e_N^r \quad y \quad \bar{\sigma}_N = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (e_N^r - \bar{e}_N)^2} \quad (5.11)$$

5.7.2 Comportamiento general de los diferentes métodos

Seguidamente se lleva a cabo un proceso de verificación y validación según el esquema de pruebas diseñado y descrito en la sección 5.4. El comportamiento de los métodos se evalúa en función del porcentaje medio de error, de modo que a menor porcentaje mejor será el comportamiento del método y viceversa. Los métodos utilizados para la comparación son, por un lado los clasificadores individuales y por otro los combinados que utilizan aquéllos como base. Dentro de los combinados, se distinguen a su vez los clásicos descritos en el capítulo tres y las estrategias locales y globales propuestas en el capítulo cuatro. A continuación se identifican los métodos de clasificación utilizados mediante sus correspondientes acrónimos, que servirán para su identificación en los gráficos y tablas de resultados que aparecen en lo que resta del capítulo:

1. Métodos de *clasificación individuales*, descritos en la sección 3.3: agrupamiento borroso (*FC*), estimador paramétrico basado en máxima verosimilitud (*BP*), estimador basado en la ventana de Parzen (*BN*), Mapas Auto-Organizativos (*SO*) y Cuantización Vectorial (Learning Vector Quantization, *L1* y *L2*) según sus dos versiones.
2. Métodos de *clasificación combinados* clásicos: Votador Mayoritario (*MV*) según la ecuación (3.45); Media aritmética (*ME*), definida en la ecuación (3.51); Mínimo (*MI*) y Máximo (*MA*), ambas de acuerdo con la ecuación (3.52), Producto (*PR*) según la ecuación (3.53). Todos ellos utilizan para la combinación los seis clasificadores individuales del punto uno.
3. Métodos *fuzzy*, son los dos métodos pertenecientes a esta categoría, a saber: *a*) el método de combinación basado en la integral fuzzy (*IF*) descrito en las secciones 3.5.2.6 y 4.3; *b*) el operador de agregación de Yager (1988) definido mediante la ecuación (3.57) e identificado como *FA*. Ambos utilizan para la combinación los seis clasificadores simples anteriores.
4. Método de combinación mediante la teoría de la decisión multi-criterio fuzzy, *MCDMF* descrito en la sección 4.2 e identificado de aquí en adelante como *MC*. Utiliza los mismos seis clasificadores individuales del punto uno.

5. Método de combinación mediante la teoría de los Mapas Cognitivos Fuzzy descritos en la sección 4.4, *MCFs* e identificado a partir de ahora como *MF*. Para la combinación utiliza sólo los clasificadores individuales *FC* y *BP*.
6. Método de relajación basado el esquema propuesto en la sección 4.4, *REL*, identificado en lo que sigue como *RL*. Para la combinación utiliza sólo los clasificadores individuales *BP* y *FC*.

Como ya sabemos, los métodos incluidos en los puntos uno a cuatro, pertenecen a la categoría de métodos locales por realizar la clasificación píxel a píxel, mientras que los incluidos en los puntos cinco y seis realizan la clasificación basándose en la vecindad, razón por la cual se les considera pertenecientes a la categoría de métodos globales.

En la tabla 5.9 se muestran los porcentajes medios y desviaciones estándar del error de clasificación para las diferentes estrategias en los tres PASOS en los que se realizan las pruebas y para los cuatro conjuntos S0, S1, S2 y S3.

Los resultados mostrados se agrupan atendiendo a las cuatro categorías que se indican a continuación:

1. Agrupación de métodos combinados locales, que utilizan técnicas de lógica fuzzy, a saber: *MC*, *IF* y *FA*.
2. Agrupación de métodos combinados globales, *MF* y *RL*, mostrándose en cada paso entre corchetes el número medio de iteraciones por imagen requeridas para lograr la convergencia. Este valor se calcula teniendo en cuenta las iteraciones empleadas por cada imagen en cada conjunto SN y promediando posteriormente entre ellos con aproximación al entero más próximo.
3. Agrupación de métodos combinados locales clásicos, incluyendo: *MV*, *MA*, *MI*, *PR* y *ME*.
4. Agrupación correspondiente a los mismos seis clasificadores individuales siguientes: *FC*, *BP*, *BN*, *SO*, *L1* y *L2*.

En cada categoría se identifica, mediante el marcado en negrita, el valor del porcentaje del menor error según la categoría y para cada uno de los cuatro conjuntos S0, S1, S2 y S3 en los tres PASOS.

Tabla 5.9 Porcentajes medios de error y desviaciones estándar para cada PASO con los cuatro conjuntos de imágenes de prueba S0, S1, S2 y S3

		PASO 1				PASO 2				PASO 3			
\tilde{e}_N : porcentaje medio de error (%)		S0		S1		S0		S2		S0		S3	
$\tilde{\sigma}_N$: desviación estándar del error (%)		\tilde{e}_0	$\tilde{\sigma}_0$	\tilde{e}_1	$\tilde{\sigma}_1$	\tilde{e}_0	$\tilde{\sigma}_0$	\tilde{e}_2	$\tilde{\sigma}_2$	\tilde{e}_0	$\tilde{\sigma}_0$	\tilde{e}_3	$\tilde{\sigma}_3$
Métodos combinados locales de naturaleza fuzzy	MC (Teoría de la decisión fuzzy)	20.6	0.90	22.1	0.81	18.2	0.88	17.6	0.88	15.1	0.85	17.5	0.91
	IF (Integral fuzzy)	20.8	0.92	22.4	0.82	18.5	0.88	17.8	0.90	15.6	0.86	17.8	0.90
	FA (Agragación fuzzy de Yager)	25.1	1.13	26.2	1.21	23.8	1.12	24.1	1.18	20.9	1.10	20.1	1.12
Métodos combinados globales (iterativos)	[número iteraciones promedio]	[16]		[18]		[14]		[15]		[11]		[12]	
	MF (Mapas Cognitivos Fuzzy)	21.7	0.98	21.6	1.02	19.2	0.96	19.9	1.11	16.1	0.93	18.5	0.96
	[número iteraciones promedio]	[19]		[22]		[18]		[18]		[14]		[15]	
	RL (Relajación)	23.2	1.09	23.6	1.10	21.1	1.05	20.9	1.12	18.4	1.03	19.1	1.02
Métodos combinados locales clásicos	MV (Votación mayoritaria)	29.3	1.29	30.1	1.31	26.3	1.25	27.8	1.28	24.8	1.29	25.2	1.24
	MA (Regla del máximo)	30.4	1.33	29.6	1.37	27.8	1.32	26.9	1.35	26.6	1.30	26.0	1.28
	MI (Regla del mínimo)	36.4	2.21	36.1	2.24	32.4	2.19	34.3	2.25	29.9	2.12	28.3	2.23
	PR (Regla del producto)	29.3	1.25	29.1	1.26	26.1	1.23	26.5	1.25	25.6	1.21	25.3	1.23
	ME (Regla de la media)	28.6	1.23	28.3	1.26	24.8	1.21	26.1	1.22	24.3	1.19	24.2	1.23
Métodos locales simples	FC (Agrupamiento borroso)	32.1	1.29	30.2	1.32	27.1	1.26	27.4	1.33	25.9	1.23	25.9	1.31
	BP (Estimador paramétrico)	30.2	1.27	29.1	1.29	26.1	1.25	26.4	1.30	25.2	1.20	24.7	1.26
	BN (Estimador no paramétrico)	33.8	1.43	34.1	1.36	32.3	1.38	31.6	1.40	27.5	1.37	27.1	1.42
	SO (Mapas auto-organizativos)	32.4	1.30	30.2	1.34	28.6	1.27	27.1	1.38	26.1	1.22	26.6	1.32
	L1 (LVQ1)	33.2	1.51	32.3	1.66	28.5	1.39	28.1	1.41	25.9	1.36	27.3	1.44
	L2 (LVQ2)	32.3	1.28	32.4	1.31	29.8	1.25	27.8	1.35	25.6	1.22	25.3	1.30

Con el fin de ilustrar gráficamente el comportamiento del conjunto testigo S0, que ha sido clasificado durante los tres PASOS señalados, a continuación se representan en los correspondientes diagramas de barras los porcentajes medios de error según las categorías de clasificadores identificados en la tabla 5.9. El clasificador con el menor porcentaje medio de error se selecciona para una representación conjunta de los mejores clasificadores según las cuatro categorías establecidas.

La figura 5.13 muestra los porcentajes medios de error para el conjunto S0 obtenidos mediante los seis clasificadores locales individuales: FC, BP, BN, SO, L1 y L2.

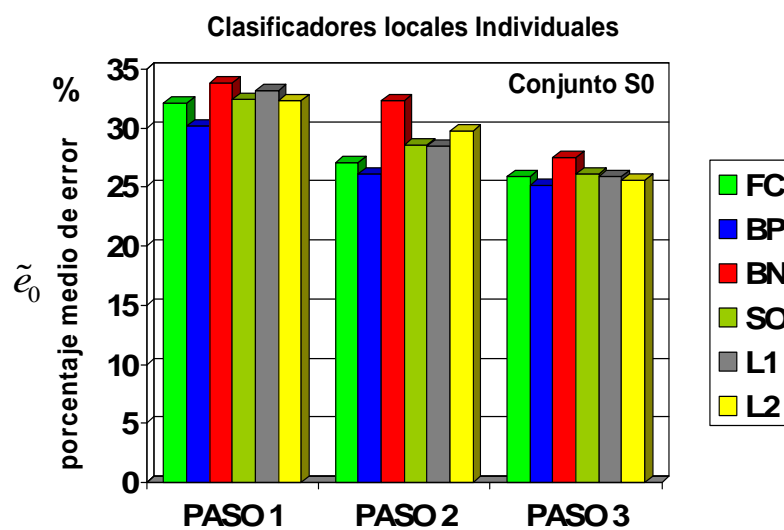


Figura 5.13 Porcentaje medio de error de los clasificadores individuales para S0 en los tres PASOS

La figura 5.14 muestra los porcentajes medios de error para el conjunto S0 obtenidos mediante los clasificadores combinados locales clásicos: MV, MA, MI, PR y ME.

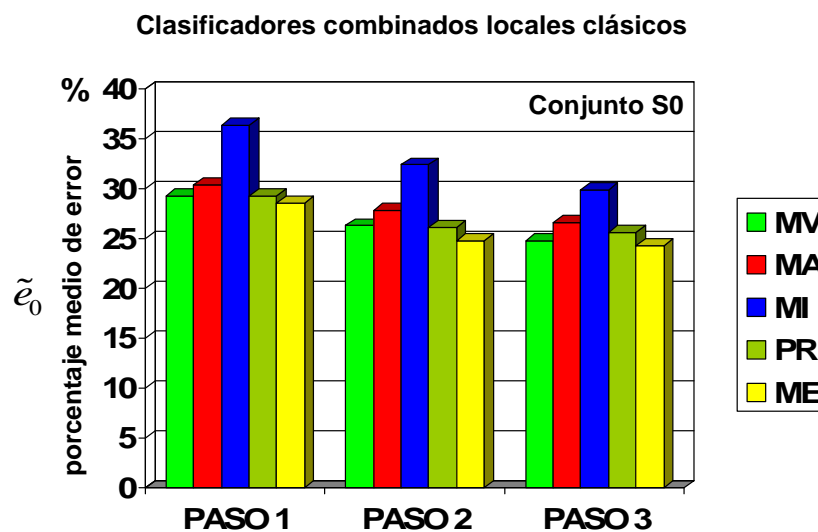


Figura 5.14 Porcentaje medio de error de los clasificadores combinados clásicos para S0 en los tres PASOS

La figura 5.15 muestra los porcentajes medios de error para el conjunto S0 obtenidos mediante los clasificadores combinados globales: MF y RL.

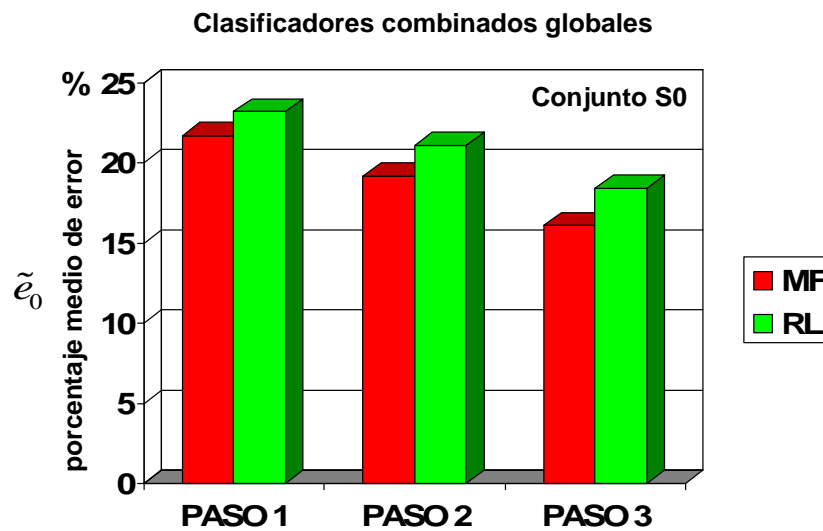


Figura 5.15 Porcentaje medio de error de los clasificadores combinados globales para el conjunto S0 en los tres PASOS

La figura 5.16 muestra los porcentajes medios de error para el conjunto S0 obtenidos mediante los clasificadores combinados locales basados en lógica fuzzy: MC, IF y FA.

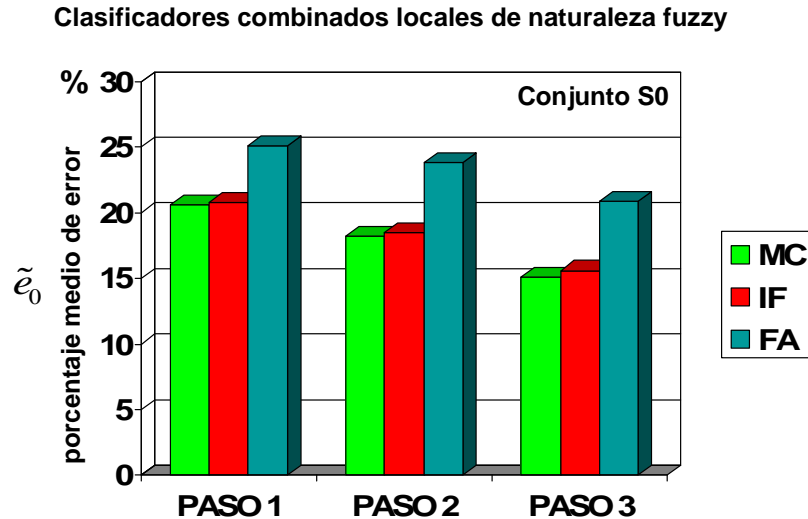


Figura 5.16 Porcentaje medio de error de los clasificadores combinados basados en lógica fuzzy para el conjunto S0 en los tres PASOS

La figura 5.17 muestra los porcentajes medios de error para el conjunto S0 obtenidos mediante los clasificadores con el mejor comportamiento dentro de cada una de las categorías, es decir: MC, MF, ME y BP.

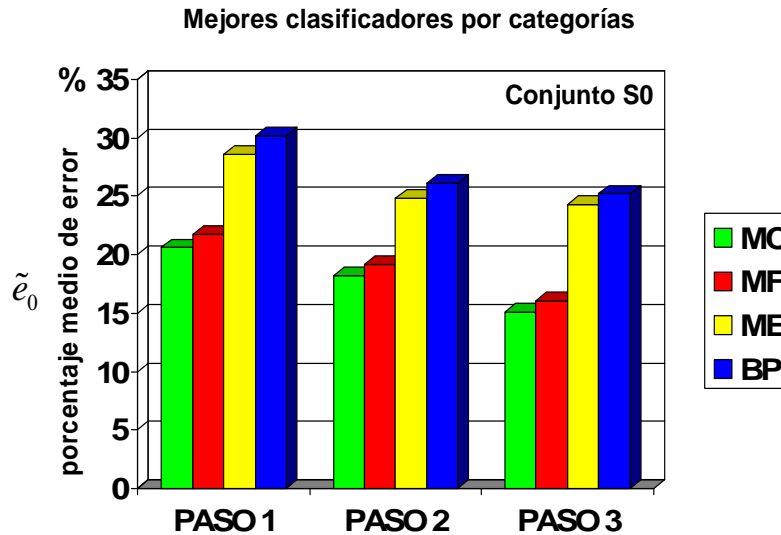


Figura 5.17 Porcentaje medio de error de los clasificadores que han proporcionado el mejor resultado dentro de su categoría para el conjunto S0 en los tres PASOS

La figura 5.18 muestra los porcentajes medios de error para los conjuntos S1, S2 y S3 en los tres PASOS, obtenidos mediante los clasificadores con el mejor comportamiento dentro de cada una de las categorías, dichos clasificadores son los mismos que en el caso de la figura 5.17: MC, MF, ME y BP.

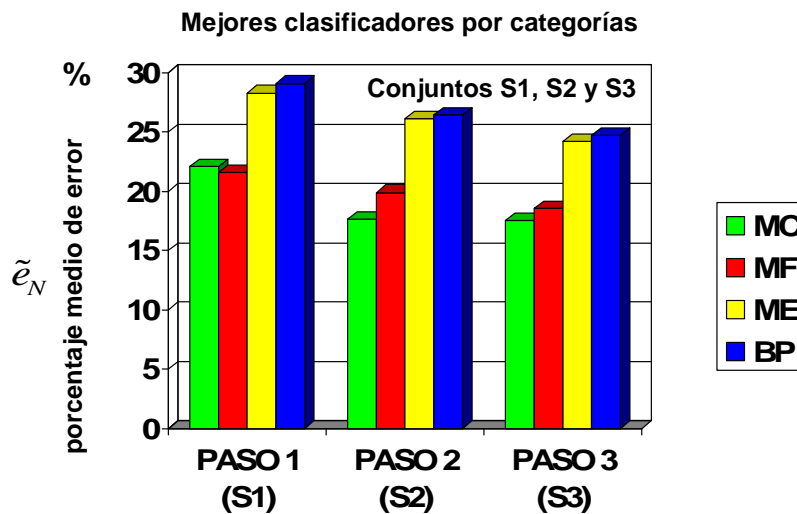


Figura 5.18 Porcentaje medio de error de los clasificadores que han proporcionado el mejor resultado dentro de su categoría para los conjuntos S1, S2 y S3 en los tres PASOS

5.7.3 Análisis de los tiempos de cómputo

5.7.3.1 Fase de decisión: métodos globales

El principal inconveniente de los métodos combinados globales es el tiempo computacional requerido para llegar a la convergencia durante la fase de decisión debido fundamentalmente a su carácter iterativo. Las pruebas se han realizado bajo un

procesador Pentium M, 1.86 *GHz* con 1 *GB* de memoria *RAM* bajo una implementación en MATLAB, tratando de optimizar el código al máximo, si bien con las implicaciones derivadas de un lenguaje de tipo interpretado como es éste. En consecuencia, los resultados obtenidos deben interpretarse desde esta perspectiva.

En la tabla 5.10 se muestran los tiempos empleados por ambos métodos globales, MF y RL, para los tres PASOS y para los conjuntos S0, S1, S2 y S3. Los números de iteraciones son los mismos que aparecen en la tabla 5.9 para sendos métodos. Los tiempos medios de ejecución por imagen se obtienen calculando la media aritmética sobre los tiempos requeridos por las diez imágenes en cada conjunto SN.

Tabla 5.10 Número medio de iteraciones y tiempos medios de ejecución por imagen en cada uno de los conjuntos S0, S1, S2 y S3 en los tres PASOS

Cómputos referidos por imagen		PASO 1		PASO 2		PASO 3	
		S0	S1	S0	S2	S0	S3
MF	Número medio de iteraciones	16	18	14	15	11	12
	Tiempo medio de ejecución (segundos)	21.6	24.4	19.9	20.8	16.6	17.8
RL	Número medio de iteraciones	19	22	18	18	14	15
	Tiempo medio de ejecución (segundos)	23.2	25.8	22.1	22.5	16.9	18.8

A partir de los resultados mostrados en la tabla 5.10 se puede calcular el tiempo medio por iteración y por imagen, teniendo en cuenta el número total de iteraciones y el tiempo total, resultando ser de 1.41 segundos en el caso del método MF y 1.22 segundos para RL.

5.7.3.2 Fase de decisión: métodos locales combinados e individuales

Todos los métodos de naturaleza local, tanto individuales como combinados, toman sus decisiones en función de las salidas proporcionadas por los clasificadores individuales. A partir de dichas salidas, los métodos combinados realizan una serie de operaciones matemáticas para generar la salida final. Por otro lado, tanto en los combinados como en los simples, las decisiones se toman píxel a píxel dada su naturaleza local. Los tiempos medios de decisión por imagen empleados por todos los clasificadores locales son del mismo orden de magnitud y además muy similares, siendo ligeramente inferiores en el caso de los clasificadores individuales por el hecho de que no realizan las operaciones de combinación requeridas por los combinados. Dichos tiempos se sitúan en torno a los

3.4 segundos en el caso de los combinados, mientras que para los individuales los tiempos se reducen aproximadamente en dos décimas de segundo. Por tanto, cabe deducir que el grueso del cómputo proviene realmente de los accesos píxel a píxel antes que de las operaciones de combinación mismas.

5.7.3.3 Fase de entrenamiento: métodos individuales

En este nivel se analizan los tiempos requeridos durante la fase de entrenamiento por los métodos individuales. No obstante, sólo se consideran significativos los tiempos empleados por los métodos iterativos, es decir *FC*, *SO*, *L1* y *L2*. Esto es así, porque *BP* consume sus tiempos en el cómputo de las medias y matrices de covarianza de las clases y *BN* solamente en la obtención de estas últimas. En ambos casos, estos tiempos resultan poco significativos si se comparan con los de los métodos iterativos, que además de iterar tienen que realizar operaciones de cálculo similares a las de *BN* o *BP*; es decir, calcular los grados de pertenencia en el caso de *FC* o los centros de las clases en el resto.

En la tabla 5.11 aparece el número de muestras, el número de iteraciones y los tiempos de ejecución en las tres fases de entrenamiento identificadas en la sección 5.4 y figura 5.4.

Tabla 5.11 Número de muestras, iteraciones y tiempos de ejecución para *FC*, *SO*, *L1* y *L2* durante la fase de entrenamiento

		Entrenamiento inicial	PASO 1	PASO 2
Número de muestras		756432	3377872	5999312
FC	Número de iteraciones	14	11	9
	Tiempo de ejecución (segundos)	12.8	31.1	51.1
SO	Número de iteraciones	15	12	10
	Tiempo de ejecución (segundos)	14.8	33.7	58.7
L1	Número de iteraciones	16	14	11
	Tiempo de ejecución (segundos)	16.8	38.3	62.1
L2	Número de iteraciones	16	15	12
	Tiempo de ejecución (segundos)	17.0	39.4	62.6

De acuerdo con los resultados de la tabla anterior, al observar las tres fases de entrenamiento (inicial, PASO 1 y PASO 2) se deduce que a mayor número de muestras, menor es el número de iteraciones requeridas para conseguir la convergencia. Sin embargo, a pesar de disminuir el número de iteraciones los tiempos de ejecución se incrementan, lo que significa que dicho incremento se debe al número de muestras procesadas.

5.7.4 Análisis comparativo de resultados

A la vista de los resultados obtenidos en las secciones 5.7.2 y 5.7.3 se pueden realizar las observaciones que se exponen a continuación según los diferentes epígrafes. Uno de los objetivos planteados en la estrategia de pruebas consistía en verificar cómo afecta el grado de aprendizaje a los resultados de la clasificación, este análisis se lleva a cabo en la sección 5.7.4.1. El segundo y principal objetivo consistía en comprobar la eficacia de los métodos combinados propuestos en este trabajo de investigación frente a otras estrategias clásicas tanto combinadas como individuales; pues bien, a la vez que se lleva a cabo esta tarea, se comprueba la eficacia de los métodos combinados clásicos frente a los métodos individuales. Por esta razón, primeramente se analizan los resultados relativos a las estrategias locales individuales en la sección 5.7.4.2, para continuar con el análisis de los métodos combinados locales clásicos en la sección 5.7.4.3, que sirven de referencia para la comparación del resto de estrategias combinadas. En las secciones 5.7.4.4 y 5.7.4.5 se analiza el comportamiento de las estrategias propuestas en este trabajo de investigación, distinguiendo entre las de naturaleza local y global respectivamente. Finalmente, en la sección 5.7.4.6 se realiza una valoración global de los resultados.

5.7.4.1 Influencia del aprendizaje

El objetivo de realizar las pruebas en tres PASOS, era precisamente para verificar la influencia del grado de aprendizaje en el proceso. Pues bien, a la vista de los resultados obtenidos, se constata que el grado de aprendizaje es determinante en los resultados. A medida que el sistema adquiere un mayor grado de aprendizaje los errores de clasificación disminuyen. Esta afirmación se comprueba con los resultados de clasificación obtenidos por el conjunto S0 en los tres PASOS. En efecto, los mejores resultados obtenidos en la clasificación de dicho conjunto se obtienen en el PASO 3,

que es exactamente donde el grado de aprendizaje es mayor, en comparación con los dos PASOS anteriores. Es preciso tener en cuenta que en cada PASO se añade un conjunto considerable de muestras de entrenamiento a las utilizadas en el proceso de entrenamiento previo. Esta mejora es progresiva, de suerte que los resultados en el PASO 2 también mejoran los obtenidos en el PASO 1. Este comportamiento es aplicable a todos los clasificadores y categorías, tanto en su versión simple como combinada, independientemente de su naturaleza local o global. Obsérvese a este respecto, cómo el porcentaje de error disminuye progresivamente desde el PASO 1 al PASO 3 a través de los resultados que se muestran tanto en la tabla 5.9 como en las figuras 5.13 a 5.17 para el conjunto testigo S0.

5.7.4.2 Clasificadores individuales

Centrándonos en los seis clasificadores individuales, *FC*, *BP*, *BN*, *SO*, *L1* y *L2*, se constata la mejora general de los resultados proporcionados por todos ellos a medida que aumenta el grado de aprendizaje. No obstante, conviene destacar que los mejores resultados se obtienen con *BP*. Esta circunstancia resulta ciertamente significativa al tratarse de un estimador clásico de naturaleza supervisada, que no realiza un proceso de aprendizaje iterativo, pudiéndose evitar así el coste computacional que marca los procesos iterativos, tal y como se ha puesto de manifiesto en la sección 5.7.3. Esto, unido al hecho de que tampoco requiere un ajuste inicial de parámetros, como se ha explicado en la sección 5.5.2 en relación al resto de métodos individuales, hace de él un candidato excelente a tener en cuenta en los procesos de clasificación, al menos en el caso de imágenes con texturas naturales. En cualquier caso, el buen comportamiento de este clasificador ha sido puesto de manifiesto en diversas aplicaciones basadas en procesos de clasificación en general (Duda y col., 2001; Escudero, 1977). Las afirmaciones previas pueden comprobarse a partir de los resultados mostrados en la tabla 5.9, así como en la figura 5.12 donde se observa cómo el porcentaje medio de error disminuye progresivamente con los tres PASOS de la prueba y además en cada PASO se produce una disminución en relación al resto de clasificadores individuales.

Atendiendo al menor porcentaje de error obtenido por los diferentes clasificadores, exceptuando *BP*, el orden queda establecido como sigue: *FC*, *SO*, *L1*, *L2* y *BN*. Según este orden, resulta de especial interés destacar el hecho de que el estimador no paramétrico basado en la ventana de Parzen (*BN*) resulta ser el que peores resultados

obtiene. Esto, según se desprende del estudio de los datos, se debe a la alta dispersión de los mismos en las clases, tal y como puede observarse en las figuras 5.6(b) y (c) para la imagen mostrada. Este comportamiento es muy similar en el resto de las imágenes estudiadas. Por este motivo, cabe deducir que en este tipo de distribuciones de datos resulta más eficaz elegir métodos que estiman un representante de las clases en lugar de las propias muestras, como ocurre con el resto de los clasificadores individuales. No obstante, dentro de éstos parece conveniente incluir en la estimación algún parámetro que de una u otra forma contemple dichas dispersiones. Justamente ambas cosas son tenidas en cuenta por el clasificador *BP* a través de la media y matrices de covarianza, lo que justificaría su buen comportamiento frente al resto de clasificadores individuales.

5.7.4.3 Clasificadores combinados locales clásicos

Refiriéndonos en este caso a *MV*, *MA*, *MI*, *PR* y *ME*. De nuevo, siguiendo con la tónica general, los resultados obtenidos con los clasificadores de este grupo mejoran progresivamente desde el PASO 1 al PASO 3. Esta mejora, sin duda, es consecuencia a su vez de las mejoras observadas en los propios clasificadores individuales, que participan en la combinación. Las combinaciones basadas en la media (*ME*) y el producto (*PR*) obtienen, por este orden, los mejores resultados dentro de esta categoría. La ligera mejoría en el caso de la primera, puede observarse tanto en la tabla 5.9 como en la figura 5.14. Respecto de la comparación entre estos dos clasificadores combinados, existe una cierta discrepancia en relación a la ligera mejoría de *ME* frente a *PR* en los resultados obtenidos en este trabajo, si se tienen en cuenta otros experimentos como los llevados a cabo por Tax y col. (2000). En efecto, en esta referencia se concluye que la combinación basada en la regla del producto obtiene mejores resultados que cualquiera de las otras combinaciones allí analizadas, incluyendo la media. No obstante, en esta misma referencia se sostiene que esto ocurre cuando los clasificadores se entrenan en espacios de características con propiedades independientes entre sí, mientras que cuando las propiedades son dependientes, el comportamiento de *ME* y *PR* es similar e incluso el primero llega a superar en eficacia al segundo. Esta conclusión es la que realmente justifica el mejor rendimiento de *ME* en el caso de las imágenes de texturas naturales utilizadas en nuestros experimentos. En efecto, en nuestro caso, las propiedades son los valores espectrales R, G y B, basadas en el mapa de color RGB, pues bien, según Littmann y Ritter (1997) o Cheng y col. (2001)

en este modelo de color existe un alta correlación entre las tres componentes espectrales, por lo que no se da la propiedad de independencia, prevista por Tax y col. (2000). Se entiende por alta correlación el hecho de que si la intensidad cambia de forma general, las tres componentes espectrales también varían proporcionalmente de acuerdo con dicho cambio de intensidad.

5.7.4.4 Clasificadores combinados globales

Son los métodos iterativos *MF* y *RL* con un comportamiento similar en términos de porcentaje de error. No obstante, *MF* es el que manifiesta un mejor comportamiento tanto en lo que se refiere al menor porcentaje medio de error según la tabla 5.9, como al menor número de iteraciones y consecuentemente menor tiempo de ejecución para la decisión, tabla 5.10. En ambos clasificadores el mayor inconveniente proviene de los elevados tiempos de ejecución para la clasificación. No obstante, este inconveniente se compensa con la mayor efectividad obtenida con cada uno de ellos frente a los clasificadores individuales, a los combinados clásicos y al *FA*. El buen comportamiento de estos dos métodos globales se debe esencialmente a dos razones principales:

1. A la utilización en la combinación de los clasificadores individuales más prometedores para la combinación, como es el caso de *BP* y *FC*.
2. A su naturaleza global, que permite considerar los principios de proximidad, similitud y conectividad de la Gestalt, involucrados en los coeficientes de regularización, ecuaciones (4.19) y (4.25) y contextual según la ecuación (4.21).

5.7.4.5 Clasificadores combinados locales basados en lógica fuzzy

En esta categoría se incluyen *MC*, *IF* y *FA*, los tres se caracterizan por: a) ser de naturaleza local, b) su concepción derivada de la lógica fuzzy y c) utilizar los seis clasificadores individuales habituales descritos en este trabajo.

Destaca el hecho de que los resultados obtenidos por *MC* e *IF* son muy similares entre ellos, siendo ligeramente mejores los obtenidos por *MC*. Sin embargo, las diferencias observadas entre los resultados de *MC* y *FA* comienzan a ser significativas. Un hecho evidente a destacar es que estas diferencias no provienen del número de clasificadores individuales utilizados para la combinación, ya que en todos los casos son seis, como se ha indicado previamente. Tampoco provienen de su naturaleza local, por el hecho de

que el procesamiento en ellos se realiza píxel a píxel. Estas dos hipótesis pueden confirmarse, en cierta medida, al comprobar los mejores resultados obtenidos por *MC*, *IF* y *FA* frente a los resultados proporcionados por los métodos combinados locales clásicos, ya que todos ellos coinciden tanto en su carácter local como en la utilización de los seis clasificadores individuales.

Descartada la localidad y el número de clasificadores individuales, el mejor comportamiento se puede achacar al propio mecanismo de combinación dentro del contexto fuzzy. En este sentido, destaca el hecho de que tanto en *MC* como en *IF* se utilizan las competencias absolutas de cada clasificador individual, sección 5.5.1, a partir de las cuales se obtienen, por un lado, los pesos asignados a cada criterio que interviene en la decisión en el caso de *MC* y por otro lado, las densidades fuzzy, que intervienen en las decisiones de *IF*. Dichas competencias no se utilizan en el caso de *FA*. Además, en este último clasificador, se da la circunstancia de que necesita un ajuste previo de su parámetro p , sección 5.3.2.2, lo cual influye en los resultados que proporciona, ello a pesar de que la estimación de dicho parámetro se haga mediante el procedimiento de validación cruzada descrito en dicha sección. Estas dos circunstancias son las que avalan el mejor comportamiento de *MC* e *IF*. Una confirmación adicional, que corrobora lo anterior, proviene del hecho de que otros operadores de agregación fuzzy que requieren la estimación de parámetros obtienen resultados similares al operador de Yager (1988). Este es el caso de la unión e intersección de Hamacher (Zimmermann, 1991) o la unión e intersección de Dubois y Prade (1982), que han sido probados en este trabajo de investigación.

5.7.4.6 Análisis global

A la vista de los resultados obtenidos se puede concluir de forma general lo siguiente:

1. Los mejores resultados se obtienen con el método *MC*, que combina convenientemente seis clasificadores individuales (*FC*, *BP*, *BN*, *SO*, *L1* y *L2*) bajo el paradigma de la teoría de la decisión multi-criterio. En esta combinación se considera la importancia relativa de los clasificadores individuales, a través de los pesos específicos asignados a los criterios, que intervienen en la decisión.
2. Otro aspecto importante a destacar, es el hecho de que a pesar de que tanto *MF* como *RL* sólo utilizan dos clasificadores (*FC* y *BP*), su comportamiento resulta

ser satisfactorio en cuanto a efectividad se refiere, hasta el punto de que se sitúan en segunda y tercera posición respectivamente dentro del conjunto de categorías. Esto es una consecuencia derivada de su naturaleza global por el hecho de integrar la información espacial, que caracteriza las imágenes, a través de los principios de la Gestalt.

3. El mayor inconveniente de los métodos *MF* y *RL* es su coste computacional, derivado de su naturaleza iterativa.
4. Lo anteriormente expuesto, es válido tanto para el conjunto *S0*, que se procesa en los tres PASOS reseñados en el proceso de pruebas, como para los conjuntos *S1*, *S2* y *S3*, figura 5.17, que se procesan en un único PASO. Se observa en dicha figura, como se ha puesto previamente de manifiesto, que los mejores resultados en términos de porcentajes medios de error, se obtienen mediante *MC*, seguido de *MF* y *RL*.
5. Los porcentajes medios de error obtenidos no suponen el óptimo, ya que en el mejor de los casos nos situamos alrededor del 15%, de donde se deduce la dificultad existente para disminuir esos porcentajes de error.

5.7.5 Ilustración de resultados relevantes

En la figura 5.19 se muestra una imagen original perteneciente al conjunto *S3*. En la figura 5.20(a)-(b) se muestran los resultados de clasificación obtenidos en el PASO 3 mediante los clasificadores *FC* y *BP*.



Figura 5.19 Imagen original perteneciente al conjunto *S3*

En la figura 5.21(a)-(b) se muestran los resultados de clasificación obtenidos por *BN* y *SO* respectivamente. No se muestran los resultados para *L1* y *L2* por no aportar ningún aspecto relevante respecto de los mostrados para *SO*.

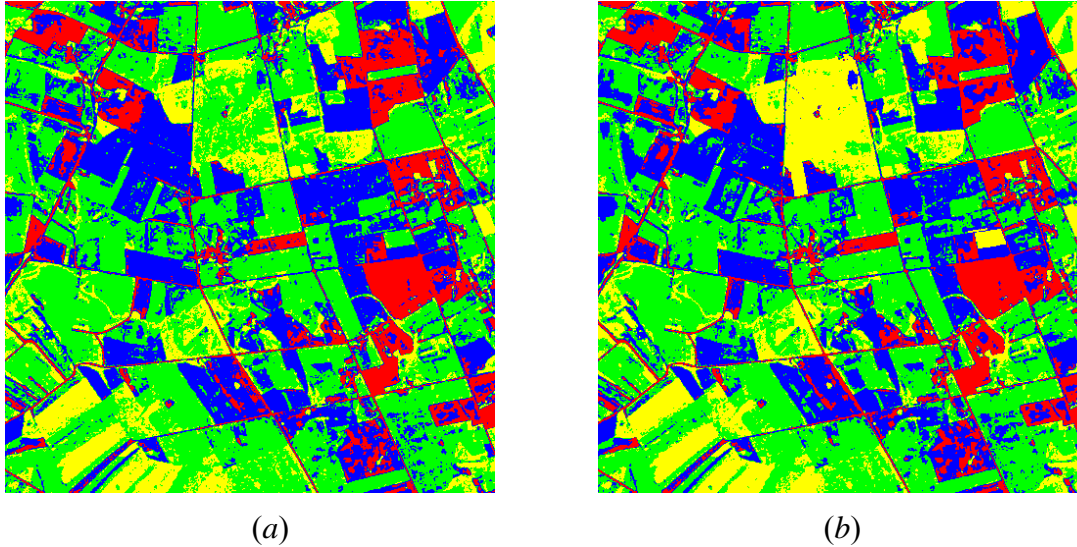


Figura 5.20 Resultados de la clasificación para la imagen de la figura 5.19 obtenidos mediante: (a) el clasificador FC y (b) el clasificador BP

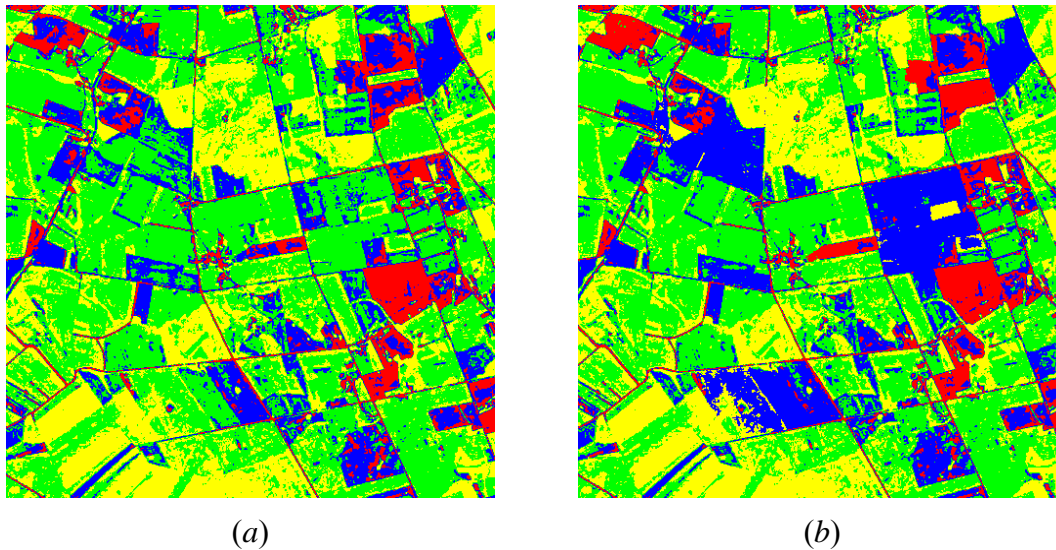


Figura 5.21 Resultados de la clasificación para la imagen de la figura 5.19 obtenidos mediante: (a) el clasificador BN y (b) el clasificador SO

Como puede apreciarse, los resultados obtenidos por el clasificador *BN*, según se visualiza en la figura 5.21(a) son los más discrepantes respecto de los resultados obtenidos con el resto de clasificadores individuales mostrados en las figuras 5.20(a) y (b) y 5.21(b). Este comportamiento es una característica general para todas las imágenes, siendo la razón que explica el peor comportamiento de *BN* frente al resto de clasificadores individuales.

En la figura 5.22(a) y (b) se muestran los resultados de clasificación correspondientes a los clasificadores *MF* y *MC* respectivamente. Los resultados obtenidos con el clasificador *RL* son similares a los obtenidos por *MF* y por tanto, no se muestran.

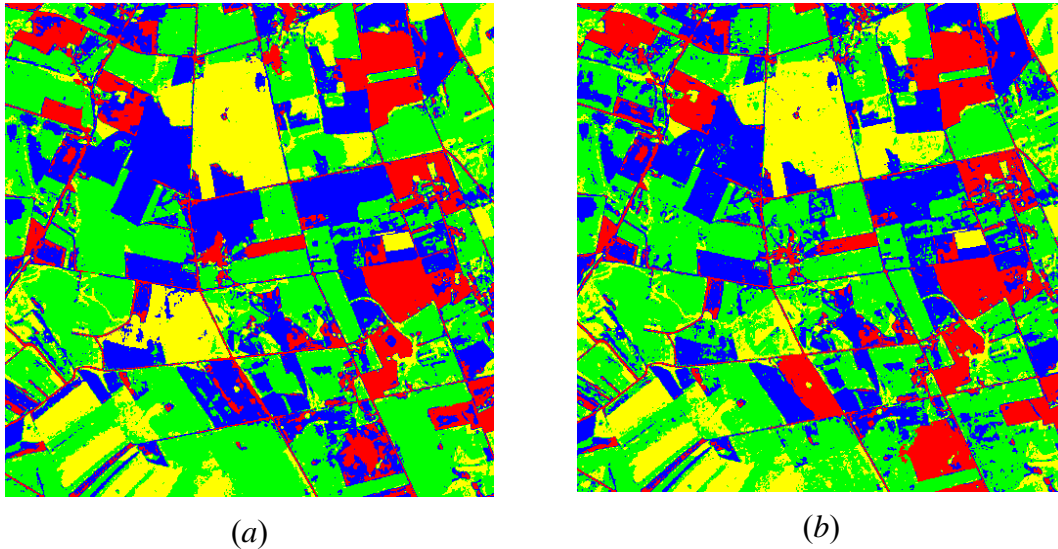


Figura 5.22 Resultados de la clasificación para la imagen de la figura 5.19 obtenidos mediante: (a) el clasificador MF y (b) el clasificador MC

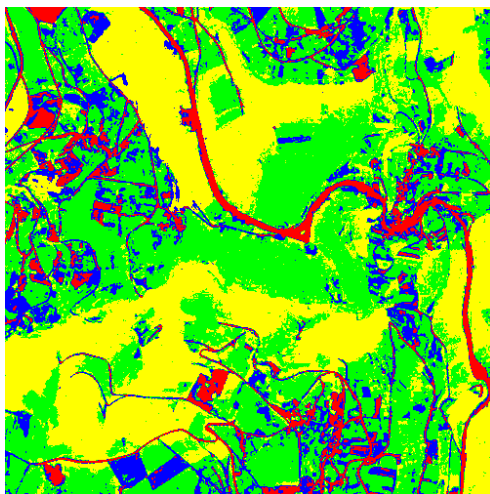
En la figura 5.22(a) se aprecia una importante homogenización en las diferentes regiones en relación a los resultados obtenidos con los métodos individuales según las figuras 5.20 y 5.21. Esto se debe al proceso global inmerso en el clasificador *MF*, de suerte que la mayoría de píxeles aislados que aparecen en determinadas regiones son absorbidos durante dicho proceso de acuerdo con el contenido de las texturas que le rodean. Obsérvese en este sentido la desaparición de puntos aislados en las grandes regiones de texturas. No obstante, en algunas situaciones, no se consigue restaurar la región completa como ocurre en el caso de las áreas en la parte inferior derecha correspondientes a la clase 4 (etiquetada de color rojo), cosa que si se logra en el caso del clasificador *MC*, como se aprecia en la imagen 5.22(b).

En el caso del clasificador *MC* el efecto de homogeneización de las regiones no se aprecia. Este hecho hace que mantenga más elementos diferenciadores dentro de las regiones. Estos elementos diferenciadores se corresponden con píxeles que están bien clasificados a pesar de estar rodeados de píxeles pertenecientes a distintas clases. Esto es justamente lo que ocurre en el caso de la región central etiquetada como perteneciente a la clase 1 (amarilla) conteniendo áreas de la clase 3, que se siguen clasificando como tal.

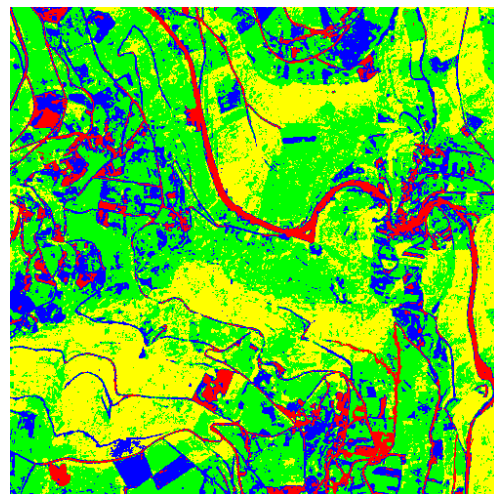
Otro efecto que aparece en el caso del clasificador *MF* es la atenuación de estructuras finas, tales como los caminos que aparecen en la imagen original mostrada en la figura 5.23, perteneciente a una imagen del conjunto S2. Obsérvese este hecho de forma destacada en la imagen 5.24(a), particularmente en la parte central, donde un número considerable de este tipo de estructuras ha desaparecido. Esta circunstancia no se aprecia en la imagen de la figura 5.24(b), donde dichas estructuras se mantienen. Por el contrario, el efecto de homogeneización de áreas extensas no se aprecia en la figura 5.24(b), mientras que en la figura 5.24(a) es evidente. Aunque quizás, estos dos últimos casos sean un tanto exagerados, no quita para que la tendencia general de estos dos clasificadores sea la expuesta. Por ello, dependiendo de la aplicación, se optará por una u otra estrategia según prime la homogeneización o la preservación de estructuras finas.



Figura 5.23 Imagen original perteneciente al conjunto S3



(a)



(b)

Figura 5.24 Resultados de la clasificación para la imagen de la figura 5.23 obtenidos mediante: (a) el clasificador MF y (b) el clasificador MC

Estas mismas observaciones, en mayor o menor medida son aplicables al resto de las imágenes procesadas a lo largo de los diferentes experimentos realizados.

5.8 Resumen del capítulo

Este capítulo resulta de especial relevancia en el sentido de que sirve para validar las estrategias combinadas propuestas en este trabajo de investigación, mediante la comparación con otros métodos de clasificación, tanto individuales como combinados clásicos.

En primer lugar, se realiza una descripción de los datos utilizados en la validación, distinguiendo entre aquellos que sirven para ajustar los parámetros iniciales requeridos por los diferentes métodos, tanto en la fase de aprendizaje como en la de decisión. En segundo lugar, se describen las imágenes con las que se realizan las pruebas de los diferentes métodos.

Se describe exhaustivamente el procedimiento de ajuste de parámetros mediante las técnicas de validación cruzada, ampliamente contrastadas en los procesos de reconocimiento de patrones en general.

Se determina el mecanismo para estimar la mejor partición sobre un conjunto de datos del que no se conoce ni el número de clases, ni por supuesto la distribución de las muestras en las clases. También se establece el procedimiento para obtener las competencias asignadas a cada clasificador, que se utilizan en las estrategias combinadas locales propuestas, concediendo distintas relevancias a los clasificadores individuales.

Mediante el diseño de la estrategia de pruebas propuesta, se definen las fases y pasos establecidos para determinar cómo afecta el grado de aprendizaje a los resultados obtenidos por los diferentes métodos.

A falta de una base de imágenes de referencia, se describe el procedimiento para su creación, con las que se obtienen los porcentajes medios de error de cada clasificador.

Finalmente, se analizan los resultados obtenidos, proporcionando tanto detalles cuantitativos como cualitativos sobre el comportamiento de los clasificadores analizados según sus categorías.

Capítulo 6

Conclusiones y trabajo futuro

6.1 Introducción

Llegados a este punto se hace necesario plantear un balance global sobre las líneas de investigación seguidas, que desembocan en las conclusiones expuestas. Por otra parte, se abren una serie de perspectivas de futuro en relación con los temas de investigación abordados y sus posibilidades de ampliación futuras.

6.2 Conclusiones

En esta sección se sintetizan las contribuciones realizadas así como las conclusiones más relevantes que se derivan de ellas. El núcleo central de la investigación ha consistido en abordar la problemática de la clasificación de texturas naturales en imágenes de exterior desde la perspectiva de la combinación de clasificadores, que por otra parte constituye una de las principales líneas de investigación abiertas actualmente en el área, tal y como se pone de manifiesto en la literatura especializada.

6.2.1 Estrategias combinadas propuestas

Se han planteado dos tipos de estrategias claramente diferenciadas, a saber: locales y globales. En cada estrategia se han propuesto, a su vez, dos métodos. En todos los métodos, independientemente de su naturaleza local o global, la combinación se lleva a cabo durante la fase de decisión, a partir de las salidas proporcionadas por los clasificadores individuales en esa misma fase. Dichos clasificadores individuales constituyen la base para la combinación.

Los dos métodos locales son: *a)* el procedimiento basado en el paradigma de la teoría de la decisión multi-criterio fuzzy, *MC* y *b)* el mecanismo basado en la integral fuzzy, *IF*. En ellos el énfasis se pone en la forma de mezclar las salidas de los clasificadores individuales, realizando la combinación píxel a píxel como corresponde a las estrategias de naturaleza local. En este sentido se establecen mecanismos apropiados para determinar las competencias de los clasificadores individuales, el objetivo consiste en otorgar a cada clasificador una cierta importancia relativa a la hora de realizar la combinación de sus salidas.

Los dos métodos globales, tienen su fundamento en la aplicación de los principios de proximidad, similitud y continuidad de la Gestalt, que llevados al campo de las imágenes permiten establecer ciertas relaciones entre un píxel y sus vecinos. Precisamente, este concepto de vecindad, siguiendo las ideas de los denominados campos aleatorios de Markov, ha permitido el diseño de los dos métodos propuestos bajo esta filosofía, a saber: *a)* los Mapas Cognitivos Fuzzy, *MF* y *b)* el método de relajación probabilista, *RL*. Como se ha indicado previamente, en ambos, la combinación se lleva a cabo en la fase de decisión a partir de las salidas proporcionadas por los clasificadores individuales durante esa misma fase, tanto para el píxel bajo clasificación como para los píxeles en la vecindad.

Una contribución adicional derivada de la investigación proviene del diseño de métodos de naturaleza no supervisada a partir de estrategias supervisadas. Aunque esto no suponga realmente una contribución novedosa, por encontrarse la idea ya recogida en la literatura especializada (Duda y col., 2001), su planteamiento específico basado en las medidas de validación de las particiones de los datos le confiere ese matiz innovador. Por otra parte, este mecanismo no supervisado ha permitido obtener las anteriormente mencionadas competencias asociadas a los clasificadores individuales.

6.2.2 Avances en la combinación de clasificadores

Los avances obtenidos mediante el diseño de los métodos de clasificación combinados se sintetizan en los puntos siguientes:

1. Una ventaja importante estriba en el hecho de que los resultados obtenidos mediante los métodos propuestos mejoran los proporcionados por los métodos combinados clásicos utilizados para la comparación.

2. Las mejoras respecto de los resultados obtenidos con los clasificadores individuales son también significativas. En este sentido, tanto este hecho como el reflejado en el punto anterior, confirman la idea generalizada en la comunidad científica, en relación a que la unión de opiniones mejora los resultados obtenidos individualmente, englobando aquí la combinación de clasificadores.
3. Se han establecido las bases para su aplicación a otros problemas que tengan que ver con temas de clasificación e identificación, incluso aunque no estén relacionados con la clasificación de texturas en imágenes. Si bien, en el caso de los métodos globales se requiere la definición del concepto de vecindad para poder aplicar los principios de la Gestalt.
4. El diseño abierto de los métodos locales, *MC* e *IF* permite su fácil ampliación para incorporar un mayor número de clasificadores y equivalentemente de criterios para la decisión.
5. Los métodos globales *MF* y *RL*, admiten igualmente, una fácil ampliación tanto en lo que respecta a la definición de nuevos coeficientes de regularización o compatibilidad como a la incorporación de nuevos clasificadores individuales al coeficiente de regularización definido en la sección 4.4.3.4.

6.3 Trabajo futuro

Es evidente, que aún queda un largo camino por recorrer ya que, como se ha indicado en la sección 5.7.4.6, el mejor porcentaje de error obtenido se sitúa alrededor del 15%. En este sentido, es posible todavía realizar un esfuerzo adicional para tratar de disminuir ese porcentaje. Aún conscientes de que lograr el óptimo, es decir el 0% de error medio, está muy lejos e incluso puede resultar inalcanzable, no es menos cierto que debe dedicarse un importante esfuerzo orientado en esa dirección.

Por este motivo, a continuación se indican algunas posibilidades de mejora de los métodos de clasificación combinados, según el tipo de estrategia local o global utilizada.

6.3.1 Métodos combinados locales

Refiriéndonos a las dos estrategias locales propuestas, a continuación se formulan algunas sugerencias de futuro.

El clasificador *MC* está basado en la teoría de la decisión, por lo que en base a sus buenos resultados, cabe pensar en la utilización de algún otro paradigma basado en dicha teoría, probablemente la teoría de la utilidad propuesta en Ríos-Insúa y col. (2002) podría ser una buena candidata.

En el caso del clasificador *IF* se pueden utilizar otros mecanismos de combinación tal como la integral fuzzy de Choquet (Grabisch, 1995 a,b ; Kuncheva, 2004).

6.3.2 Métodos combinados globales

Centrándonos en los métodos globales *MF* y *RL*, a continuación se exponen una serie de posibles mejoras, que pueden ser consideradas en futuros trabajos:

1. Se puede incluir un mayor número de clasificadores, definiendo coeficientes de regularización adicionales en función del número de clasificadores considerados.
2. En los métodos propuestos se actualizan los estados de los nodos, así en el caso de *MF* la actualización se hace sobre los grados de pertenencia proporcionados por el clasificador individual *FC*, manteniendo fijas las probabilidades generadas por el también clasificador individual *BP*. En el caso del método *RL* ocurre justo lo contrario, de suerte que se actualizan los estados a partir de las probabilidades proporcionadas por *BP*, manteniendo fijos los grados de pertenencia proporcionados por *FC*. Pues bien, una propuesta de trabajo futuro consiste en la actualización de ambos simultáneamente, de forma que en cada iteración se utilicen los dos valores actualizados.
3. Ampliar las regiones de vecindad a regiones m -conectadas, donde m tome el valor de 24, que corresponde a regiones de vecindad de 5×5 , o incluso mayores, para determinar el efecto de influencia de la vecindad según su tamaño.
4. En lugar de la tangente hiperbólica utilizada en ambos métodos globales como elemento no lineal de activación de los nodos, se pueden considerar otras funciones de activación definidas en el contexto de las redes neuronales, tales como la función logística o sigmoideal (Haykin, 1992).
5. Probar otros procesos de actualización basados en optimización, tales como *enfriamiento simulado* o el modelo de *red neuronal* de Hopfield. En ambos casos,

con el fin de controlar el proceso de convergencia a través de las correspondientes funciones de energía, que incorporan sendos procesos (Duda y col., 2001).

6.3.3 Ajuste de parámetros

Durante el proceso de validación cruzada, para el ajuste de parámetros de los métodos individuales, descritos en la sección 5.3, en lugar de utilizar un 90% de las muestras como conjunto de entrenamiento y el 10% restante como conjunto de validación, una posible mejora puede provenir por la utilización de la estrategia conocida en la literatura inglesa como “*n-fold cross validation*” (Duda y col., 2001) la cual divide el conjunto de los datos en n subconjuntos. De esta forma, el entrenamiento podría iniciarse con $n - 1$ de esos subconjuntos y la clasificación con el subconjunto restante, realizando este proceso de forma iterativa n veces de tal modo que cada subconjunto se clasifica una vez.

6.3.4 General

Se propone experimentar con otras estrategias combinadas de nivel superior, de forma que las salidas proporcionadas por cualquiera de los métodos combinados propuestos en este trabajo constituyan las entradas de otros clasificadores, tales como redes neuronales, que requieren un proceso de aprendizaje previo. Se trata de utilizar estos clasificadores en un modelo de orden superior similar al propuesto por Kuncheva (2004).

Desde el ámbito específico de la clasificación de texturas en imágenes naturales, podría resultar interesante aprovechar todo tipo de información disponible, como puede ser la proporcionada por los Mapas Digitales del Terreno cuando estén disponibles, de esa forma se podrían distinguir determinados tipos de texturas por su ubicación en el terreno según las alturas sobre el nivel del mar, por ejemplo zonas de pradera o arbustos, que pueden distinguirse por sus diferentes altitudes.

Se contempla la posibilidad de utilizar información estructural en las imágenes, de forma que, según el tipo de objetos de la escena, permita discernir entre texturas ambiguas; por ejemplo un cierto camino rural, que podría confundirse por su tonalidad ocre con una parcela de barbecho con tonalidades similares, cabe distinguirlo, por su forma alargada y estrecha, de las parcelas, que no poseen esa geometría.

Bibliografia

1. Aiyyer, A., Pyun, K., Huang, Y., O'Brien, D.B. and Gray, R.M. (2005). Lloyd Clustering of Gauss mixture models for image compression and classification. *Signal Processing: Image Communication*, 20, 459-485.
2. Allili, M.S. and Ziou, D. (2007). Globally adaptive region information for automatic color texture image segmentation. *Pattern Recognition Letters*, 28, 1946-1956.
3. Andrey, P. and Tarroux, P. (1998). Unsupervised Segmentation of Markov Random Field Modeled Textured Images Using Selectionist Relaxation. *DOI Bookmark*, 20(3), 252-262. <http://doi.ieeecomputersociety.org/10.1109/34.667883>
4. Arivazhagan, S. and Ganesan, L. (2003). Texture classification using wavelet transform. *Pattern Recognition Letters*, 24, 1513-1521.
5. Asuncion, A. and Newman, D.J. (2008). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science, available on-line <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
6. Atkinson, P.M. and Tatnall, R.L. (1997). Neural Networks in Remote Sensing. *Int. Journal. Remote Sensing*,
7. AVIRIS: Airborne Visible InfraRed Imaging Spectrometer (2009) <http://aviris.jpl.nasa.gov/html/aviris.overview.html>
8. Axelrod, R. (1976). *Structure of decision: the cognitive maps of political elites*. New York: Princeton.
9. Balasko, B., Abonyi, J. and Feil, B. (2006). Fuzzy Clustering and Data Analysis Toolbox for Use with Matlab, Veszprem University, Hungary
[URL: http://www.fmt.vein.hu/softcomp/fclusttoolbox/FuzzyClusteringToolbox.pdf](http://www.fmt.vein.hu/softcomp/fclusttoolbox/FuzzyClusteringToolbox.pdf)
10. Barnsley, M.J. and Barr, S.L. (1996). Inferring urban land use from satellite sensor images using kernel-based spatial reclassification. *Photogramm. Engineering Remote Sensing*, 62(8), 949-958.
11. Barker, S.A. and Rayner, P.J.W. (2000). Unsupervised image segmentation using Markov random field models. *Pattern Recognition* 33 (2000) 587,602.
12. Basler (2009). Basler Vision Technologies. <http://www.baslerweb.com/>
13. Battiti, R. and Colla, A.M. (1994). Democracy in neural nets: voting schemes for classification. *Neural Networks*, 7, 691-707.
14. Benediktsson, J.A. and I. Kanellopoulos, I. (1999). Classification of multisource and hyperspectral data based on decision fusion. *IEEE Trans. Geoscience Remote Sensing*, 37(3), 1367-1377.
15. Benediktsson, J.A., Swain, P. H. and Ersoy, O. K. (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Trans. Geosci. Remote Sens.*, 28(4), 540-552.
16. Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer, Plenum Press, New York.
17. Bogataj, M. and Usenik, J. (2005). Fuzzy approach to the spatial games in the total market area. *International Journal of Production Economics*, 93-94(8), 493-503.

18. Borges, J. S., Bioucas-Dias, J.M. and Marça, A.R.S. (2007). Bayesian Hyperspectral Image Segmentation with Discriminative Class Learning. (J. Martí, J.M. Benedí, A.M. Mendoça and J. Serrat (Eds.). Pattern Recognition and Image Analysis. Lecture Notes in Computer Science, Springer-Verlag, Berlin, 4477, Part I, 22-29.
19. Botte-Lecocq, C., Hammouch, K., Moussa, A., Postaire, J.G. Sbihi, A. and Touzani, A. (2007). Image Processing Techniques for Unsupervised Pattern Classification. IN Scene Reconstruction, Pose Estimation and Tracking (Rustaam Stolkin, Ed.). I-Tech, Vienna, Austria, Capter 25, 467-488.
20. Briem, G.J., Benediktsson, J.A. and Sveinsson, J.R. (2002). Multiple Classifiers applied to multisource Remote Sensing Data. IEEE Trans. Geosci. Remote Sens., 40(10), 2291–2299.
21. Brodatz, P. (1966). Textures: A photographic album for artists and designers. Dover Publications, New York.
22. Broek, E. L. and Rikxoort, E. M. (2004). Evaluation of color representation for texture analysis. In Proceedings of the 16th Belgium-Netherlands Artificial Intelligence Conference, R. Verbrugge, N. Taatgen, and L. R. B. Schomaker, (Eds.), Groningen - The Netherlands, 35–42.
23. Bruzzone, L., Fernández-Prieto, D. and Serpico, S. B. (1999). A neural-statistical approach to multitemporal and multisource remote-sensing image classification. IEEE Trans. Geosci. Remote Sens., 37(11), 1350–1359.
24. Buchanan, B. G. and Shorliffe, E.H. (Eds.) (1984). Rule-Based Expert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project, Addison-Wesley, Reading, MA.
25. Bueno, S. and Salmeron, J.L. (2009). Benchmarking main activation functions in fuzzy cognitive maps2. Expert Systems with Applications, 36(3), 5221-5229.
26. Campadelli, P., Medici, D. and Schettini, R. (1997). Color image segmentation using Hopfield Networks. Image And Vision Computing, 15, 161-166.
27. Camps-Valls, G. and Bruzzone, L. (2005) Kernel-based methods for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens., 43(6), 1351–1362.
28. Camps-Valls, G., Gómez-Chova, L., Calpe, J., Soria, E., Martín, J. D., Alonso, L. and Moreno, J. (2004) Robust support vector method for hyperspectral data classification and knowledge discovery. IEEE Trans. Geoscience. Remote Sensing, 42(7), 1530–1542.
29. Canty, M. J. (2006). Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL, CRC/Taylor & Francis.
30. Cao, J., Shridhar, M. and Ahmadi, M. (1995). Fusion of Classifiers with Fuzzy Integrals. In: Proc. 3th Int. conf. Document Analysis and Recognition (ICDAR'95), 1, (108-111)
31. Cariou, C. and Chehdi, K. (2008). Unsupervised texture segmentation/classification using 2-D autoregressive modelling and the stochastic expectation-maximization algorithm. Pattern Recognition Letters, 29, 905-917.
32. Chamorro-Martínez, J., Galán-Perales, E., Prados-Suárez B. and Soto-Hidalgo, J.M. (2007). Perceptually-Based Functions for Coariness Textural Feature Representation. (J. Martí, J.M. Benedí, A.M. Mendoça and J. Serrat (Eds.)). Pattern Recognition and Image Analysis. Lecture Notes in Computer Science, Springer-Verlag, Berlin, 4477, Part I, 578-586.
33. Chan, J.C.W., Laporte, N. and Defries, R.S. (2003). Texture Classification of logged forest in tropical Africa using machine-learning algorithms. Int. J. Remote Sensing, 24(6), 1401-1407.
34. Chang, M.M., Sezan, M.I. and Tekalp, A.M. (1994). Adaptive Bayesian segmentation of color images. Journal Electronic Imaging, 3, 404-414.

35. Chang, T. and Kuo, C.C.J. (1993). Texture analysis and classification with tree structured wavelet transform. *IEEE Trans. Image Processing*, 2(10), 429-441.
36. Chanussot, J., Benediktsson, J.A. and Fauvel, M. (2006). Classification of Remote Sensing Images from Urban Areas using a Fuzzy Possibilistic Model. *IEEE Geoscience and Remote Sensing Letters*, 3(1), 40-1407.
37. Chen, C.T. (2000). Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets and Systems* 114, 1-9.
38. Cheng, H.D., Jiang, X.H., Sun, Y. and Wang, J., (2001). Color image segmentation: advances and prospects. *Pattern Recognition* 34 (12), 2259–2281.
39. Cheng, H.D., Jiang, X.H., Sun, Y. and Wang, J., (2002). Color image segmentation based on homogram thresholding and region merging. *Pattern Recognition* 35 (2), 373-393.
40. Chen, J., Pappas, T.N., Mojsilovic, A. and Rogowitz, B. E. (2005). Adaptive Perceptual Color-Texture Image Segmentation. *IEEE Trans. Image Processing*, 14(10), 1524-1536.
41. Cho, S.B. and Kim, J.K. (1995a). Combining multiple neural networks by fuzzy integral and robust classification. *IEEE Trans. Systems, Man and Cybernetics*, 25, 380-384.
42. Cho, S.B. and Kim, J.K. (1995b). Multiple network fusion using fuzzy logic. *IEEE Trans. Neural Networks*, 6, 497-501.
43. Clark, A.A., Thomas, B.T., Campbell, N.W. and Greenway, P. (1999). Texture deconvolution for Fourier-based analysis of non-rectangular regions. *BMVC*, 193-202.
44. Comaniciu, D. and Meer, P. (1997). Robust analysis of feature spaces: color image segmentation. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Juan, PR, 750-755.
45. Comaniciu, D. and Meer, P. (2002). Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2002), 603–619.
46. Debeir, O., van den Steen, I., Latinne, P., van Ham, P. and Wolff, E. (2002). Textural and contextual land-cover classification using single and multiple classifier systems. *Photogramm. Eng. Remote Sens.*, 68(6), 597–605.
47. Deng, H. and Clausi, D.A. (2004). Unsupervised image segmentation using a simple MRF model with a new implementation scheme. *Pattern Recognition*. 37, no. 12, 2323–2335.
48. Deng, H. and Clausi, D.A. (2005a) Unsupervised segmentation of synthetic aperture radar ice imagery using a novel Markov random field model. *IEEE Trans. Geosci. Remote Sensing* 43 (3), 528–538.
49. Deng, H. and Clausi, D. A. (2005b). Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov model. *Lecture Notes in computer Science*, Calgari, Italy, 1857, 1-15.
50. Dietterich, T.G. (2000). Ensemble methods in machine learning. In J. Kittler and F. Roli (Eds.), *Multiple Classifier Systems*, 1857. *IEEE Trans. Geoscience Remote Sensing*, 43, 528-538.
51. Dimap (2009). Digital Image Processing, S.L. <http://www.dimap.es/>
52. Domingos, P. and Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29, 103–130
53. Dubois, D. and Prade, H. (1982). A class of fuzzy measures based on fuzzy triangular norms. *International Journal of Gen Systems*, 8, 43-61.

54. Dubois, D. and Prade, H. (1985). A review of fuzzy set aggregation connectives. *Information Sciences*, 36, 85-521.
55. Duda, R.O., Hart, P.E. and Stork, D.S. (2001). *Pattern Classification*. New York: Wiley & Sons.
56. Duin, R.P.W. (1976). On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE Trans. Computers*, C-25(11), 1175-1179.
57. Dunn, D. and Higgins, W.E. (1995). Optimal Gabor filters for texture segmentation. *IEEE Trans. Image Processing*, 4(7), 947-964.
58. Du, Y., Chang C, Thouin, P.D. (2004). Unsupervised approach to color video thresholding, *Optical Engineering*, 43, No. 2, 282-289
59. E-Cognition-Definiens (2009) <http://www.definiens.com/>
60. Envi (2009) <http://www.ittvis.com/ProductServices/ENVI.aspx>
61. Erdas Imagine (2009) <http://www.erdas.com>
62. Escalera, A. (2001). *Visión por Computador: Fundamentos y Métodos*, Prentice-Hall, Madrid.
63. Escudero, L.F. (1977). *Reconocimiento de patrones*. Paraninfo, Madrid
64. Fauvel, M., Chanussot, J. and Benediktsson, J.A. (2006). Decision Fusion for the Classification of Urban Remote Sensing Images. *IEEE Trans. Geoscience and Remote Sensing*, 44(10), 2828- 2838.
65. Ferreiro-Armán, M., Bandeira, L.P.C., Martín-Herrero, J. and Pina, P. (2007). Classifiers for vegetation and forest mapping with low resolution multiespectral imagery. (J. Martí, J.M. Benedí, A.M. Mendoça and J. Serrat (Eds.)). *Pattern Recognition and Image Analysis. Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 4477, Part I, 177-184.
66. Frate, F. Del, Pacifici, F., Schiavon, G. and Solimini, C. (2007). Use of Neural Networks for Automatic Classification from High-Resolution Images. *IEEE Trans. Geoscience and Remote Sensing*, 45(4), 800- 809.
67. Frizzelle, B.G. and Moody, A. (2001). Mapping continuous distributions of land cover: A comparison of maximum-likelihood estimation and artificial neural networks. *Photogramm. Eng. Remote Sens.*, 67(6), 693–705.
68. Fukunaga, K. and Hummels, M.D. (1987). Bayes error estimation using Parzen and K-NN procedures. *IEEE Trans Pattern Anal. Machine Intelligence*, 9(5), 634-643.
69. Geman, S. and Geman, D. (1984). Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis Machine Intelligence*, 6, 721-741.
70. Giacinto, G., Roli, F. and Bruzzone, L. (2000). Combination of neural and statistical algorithms for supervised classification of remote-sensing image. *Pattern Recognition Letters*, 21(5), 385-397.
71. Gong, P. and Howarth, P.J. (1990). The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. *Photogramm. Eng. Remote Sens.*, 56(1), 67–73, 1990.
72. Gonzalez, R. and Woods, R. (2007). *Digital Image Processing*, Prentice-Hall.
73. Grabisch, M. (1995). On equivalence classes of fuzzy connectives- the cause of fuzzy integrals. *IEEE Trans. Fuzzy Systems*, 3(1), 96-109.
74. Grabisch, M., Nguen, H.T. and Walker, E.A. (1995). *Fundamentals of Uncertainty Calculi, With applications to Fuzzy Inference*. Kluwer Academic Publishers, Dordrecht.

75. Grim, J., Kittler, J., Pudil, P. and Somol, P. (2002). Multiple Classifier Fusion in Probabilistic Neural Networks. *Pattern Analysis and Applications*, 5(2), 221-233.
76. Gu, X. and Zhu, Q. (2006). Fuzzy multi-attribute decision-making method based on eigenvector of fuzzy attribute evaluation space. *Decision Support Systems*, 41, 400-410.
77. Guijarro, M. and Pajares, G. (2009). On combining classifiers through a fuzzy multicriteria decision making approach: applied to natural textured images. *Expert Systems with Applications*, 36(3P2), 7262-7269.
78. Guijarro, M., Pajares, G. and Herrera, P.J. (2008). On combining classifiers by relaxation for natural images in images. *Innovations in Hybrid Intelligent Systems (HAIS08). Advances in Soft Computing (Corchado, E., Corchado, J.M. and Abraham, A. Eds.), Lecture Notes in Artificial Intelligence, Springer-Verlag Berlin Heidelberg, September, 5271, 345-352, doi: 10.1007/978-3-540-87656-4_43.*
79. Guijarro, M., Pajares, G., Abreu, R., Garmendia, L. and Santos, M. (2007a). Design of a Hybrid Classifier for Natural Textures in Images from the Bayesian and Fuzzy Paradigms. In *Proc. IEEE International Symposium on Intelligent Signal Processing (WISP07). In Conference Proceedings Book (J. Ureña, J.J. García, Eds.), 431-436, Alcalá de Henares, Madrid, October 3-5, doi: 10.1109/WISP.2007.4447562*
80. Guijarro, M., Abreu, R. and Pajares, G. (2007b). A New Unsupervised Hybrid Classifier for Natural Textures in Images. *Innovations in Hybrid Intelligent Systems (HAIS07). Advances In Soft Computing. Lecture Notes in Artificial Intelligence (Eds. Emilio Corchado, Juan M. Corchado, Ajith Abraham), 44, 280-287, Springer-Verlag Berlin Heidelberg, November, doi:10.1007/978-3-540-74972-1_37.*
81. Guijarro, M., Abreu, R. and Pajares, G. (2007c). On combining Learning Vector Quantization and the Bayesian classifiers for natural textured images. *Proc. II Congreso Español de Informática. V Taller Nacional de Minería de Datos y Aprendizaje, (TAMIDA2007), 195-201, Zaragoza, Spain, September.*
82. Haapanen, R., Ek, A.R., Bauer, M.E. and Findley, A.O. (2004). Remote Sensing Environment, 89(3), 265-271.
83. Hanmandlu, M., Madasu, V.K. and Vasikarla, S. (2004). A Fuzzy Approach to Texture Segmentation, in: *Proc. of the IEEE International Conference on Information Technology: Coding and Computing (ITCC'04), The Orleans, Las Vegas, Nevada, USA (636-642).*
84. Hansen, M, Dubayah, R. and Defries, R. (1996). Classification trees: an alternative to traditional land cover classifiers. *Int. J. Remote Sensing*, 17(5), 1075-1081.
85. Hansen, L.K. and Salamon, P. (1990). Neural Network ensembles. *IEEE Trans. Pattern Analysis and Applications*, 12(10), 993-1001.
86. Hashem, S. and B. Schmeiser (1995). Improving Model Accuracy Using Optimal Linear Combinations of Trained Neural Networks, *IEEE Transactions on Neural Networks*, 6(3), 792-794.
87. Haykin, S. (1994). *Neural Networks: a comprehensive foundation*. Macmillan College Publishing Co., New York.
88. Ho, T.K., Hull, J.J. and Srihari, S.N. (1994). Decision combination in multiple classifier systems. *IEEE Trans. Pattern Analysis and Machine Intell.*, 16, 66-75.
89. Ho, T. K. (2002). Multiple classifier combination: Lessons and the next steps. In A. Kandel and H. Bunke (Eds.), *Hybrid Methods in Pattern Recognition*. World Scientific Publishing, 171-198.
90. Huang, C.L. (1999). Pattern image segmentation using modified Hopfield model. *Pattern Recognition Letters*, 13, 345-353.

91. Huang, H.J. and Hsu, C.N. (2002). Bayesian classification for data from the same unknown class. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 2002, 32(2), 137-145.
92. Hu, R. and Damper, R.I. (2008). A "Non Panacea Theorem" for classifier combination. *Pattern Recognition*, 41(8), 2665-2673.
93. Hummel, R. and Zucker, S. (1983). On the foundations of relaxation labeling process. *IEEE Trans. Pattern Anal. Mach. Intell.*, 5, 267-287
94. Hwang, C.L. and Yoon, K. (1981). *Multiple Attributes Decision Making Methods and Applications*, Springer, Berlin Heidelberg.
95. Indra Sistemas, S.A. (2009) www.indra.es
96. Intergraph (2009) www.intergraph.com
97. ISCAR (2006). Ingeniería, Control, Automática y Robótica), aprobado por la Universidad Complutense de Madrid con número de referencia 910793. <http://www.ucm.es/info/dacya/SP/ResearchGroups.htm>
98. Jain, A.K. and Ramaswami, M.D. (1988). Classifier design with Parzen windows. In *Pattern Recognition and Artificial Intelligence toward an integration*, E.S. Geselma and L.N. Kanal (Eds.) Amsterdam, The Netherlands, 211-218.
99. Jiji, G.W. and Ganesan, L. (2008). A new approach for unsupervised segmentation. *Applied Soft Computing Journal*, doi:10.1016/j.asoc.2008.06.002.
100. Jimenez, L., Morales-Morell, A. and Creus, A. (1999). Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks. *IEEE Trans. Geosci. Remote Sens.*, 37(3), 1360-1366.
101. Kanellopoulos, I., Wilkinson, G.G. and Megier, J. (1993). Integration of neural network and statistical image classifier for land cover mapping. *Proc. Geoscience and Remote Sensing Symposium (IGARSS'93). Better Understanding of Earth Environment*, 2(18-21), 511-513.
102. Kaufmann, A. and Gupta, M.M. (1985). *Introduction to Fuzzy Arithmetic: Theory and Applications*, Van Nostrand Reinhold, New York.
103. Kim, J., Fisher III, J.W., Yezzi, A. Çetin, M. and Willsky, A.S. (2006). A non parametric statistical method for image segmentation using information theory and curve evolution. *IEEE Trans. Image Processing*, 14(10), 1486-1502.
104. Kim, D.W., Lee, K.H. and Lee, D (2003). Fuzzy Cluster validation index based on inter-cluster proximity. *Pattern Recognition Letters*, 24 (2003) 2561-2574.
105. Kittler, J., Hatef, M. and Duin, R.P.W. (1996). Combining Classifiers. *Proc. 13th. Int. Conf. on Pattern Recognition*, 2, 897-901.
106. Kittler, J., Hatef, M., Duin, R.P.W. and Matas, J. (1998). On Combining Classifiers, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20 (3), 226-239.
107. Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt.
108. Kohonen, T. (1989). *Self-Organization and Associative Memory*. Springer-Verlag, Berlin.
109. Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464-1480.
110. Kohonen, T. (1997). *Self-Organizing Maps*. Springer, Berlin.
111. Koschan, A. and Abidi, M. (2008). *Digital Color Image Processing*. Wiley & Sons, Hoboken, NJ.
112. Kosko, B. (1992). *Neural Networks and Fuzzy Systems*. Prentice-Hall, Englewood Cliffs, NJ.
113. Kosko, B. (1986). Fuzzy Cognitive Maps (1986). *Int. J. Man Machine Studies*. 24, 65-75.

114. Koulouriotis, D.E., Diakoulakis, I.E., Emiris, D.M. and Zopoundis, C.D. (2005). Development of dynamic cognitive networks as complex systems approximators: validation in financial time series. *Applied Soft Computing*, 5(2), 157-179.
115. Krishnamachari, S. and Chellappa, R. (1997). Multiresolution Gauss-Markov random field models for texture segmentation. *Image Processing, IEEE Transactions on*, 6, Issue: 2, 251-267
116. Krishnapuram, R. and Freg, C.P. (1992). Fitting an unknown number of lines and planes to image data through compatible cluster merging. *Pattern Recognition*, 25, 385-400.
117. Kumar, A.S. and Majumder, K.L. (2001). Information fusion in tree classifiers. *Int. J. Remote Sensing*, 22(5), 861-869.
118. Kumar, S., Ghosh, J. and Crawford, M.M. (2001). Best-bases feature extraction for pairwise classification of hyperspectral data. *IEEE Trans. Geoscience and Remote Sensing*, 39(7), 1368-1379.
119. Kumar, S., Ghosh, J. and Crawford, M.M. (2002). Hierarchical Fusion of Multiple Classifiers for Hyperspectral Data Analysis. *Pattern Analysis and Applications*, 5, 210-220.
120. Kuncheva, L.I., Bezdek, J.C. and Duin, R.P. (2001). Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition*, 34(2), 299-314.
121. Kuncheva, L.I. (2002). Switching between selection and fusion in combining classifiers: an experiment. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 2002, 32(2), 146-156.
122. Kuncheva, L.I. (2003). "Fuzzy" vs "non-fuzzy" in combining classifiers designed by boosting. *IEEE Trans. Fuzzy Systems*, 11, 729-741.
123. Kuncheva, L.I. (2004). *Combining Pattern Classifiers: Methods and Algorithms*, Wiley, 2004.
124. Kupinski, M.A. and Gigger, M. L. (1998). Automated seeded lesion segmentation on digital mammograms. *IEEE Trans. Medical Imaging*, 17, 510-517.
125. Lam, L. and Suen, C.Y. (1997). Application of majority voting to pattern recognition: an analysis of its behaviour and performance. *IEEE Trans. Systems, Man and Cybernetics*, 27, no. 5, 553-568.
126. LANDSAT Thematic Mapper (2009) http://edc.usgs.gov/guides/landsat_tm.html.
127. Lepisto, L, Kunttu, I., Autio, J. and Visa, A. (2003). Classification of non-homogenous texture images by combining classifiers. In *Proc. IEEE Int. Conf. Image Processing*, I, 981-984, Barcelona, Spain.
128. Lepisto, L, Kunttu, I. Autio, J. and Visa, A. (2005). Classification of Natural Images using supervised and Unsupervised Classifier Combinations. F. Roli and S. Vitulano (Eds.). *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 3617, 770-777.
129. Liew, A.W.C., Leung, S.H. and Lau, W.H. (2003). Segmentation of color images by spatial fuzzy clustering. *IEEE Trans. Fuzzy Systems*, 11, 542-549.
130. Li, H., Liu, G. and Zhang, Z. (2006). A new texture generation model based on Pseudo-DCT coefficients. *IEEE Trans. Image Processing*, 15(5), 1300-1312.
131. Li, J., Najmi, A. and Gray, R.M. (2000). Image classification by a two-dimensional hidden Markov model. *IEEE Trans. Signal Processing*, 48(2), 517-533.
132. Lillo, A., Motta, G. and Storer, J.A. (2007). Supervised segmentation based on texture signatures extracted in the Frequency Domain. (J. Martí, J.M. Benedí, A.M. Mendoça and J. Serrat (Eds.)). *Pattern Recognition and Image Analysis. Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 4477, Part I, 89-96.

133. Lim, Y.W. and Lee, S.U. (1990). On the color image segmentation algorithm based on the thresholding and the fuzzy c-means technique. *Pattern Recognition*, 23(9), 935-952.
134. Lin, X., Yacoub, S., Burns, J. and Simske, S. (2003). Performance analysis of pattern classifier combination by plurality voting. *Pattern Recognition Letters*, 24, no. 12, 1959-1969.
135. Li, S.Z. (2000). Modelling image analysis problems using Markov Random Fields. In Rao, C.R., Shanbhag, D.N. (Eds.), *Stochastic Process: Modelling and Simulation, Handbook of Statistics*, 20, 1-43.
136. Littmann, E. and Ritter, H., (1997). Adaptive color segmentation -A comparison of neural and statistical methods. *IEEE Trans. Neural Networks*, 8(1), 175-185.
137. Lo, Y.S. and Pei, S.C. (1999). Color image segmentation using local histogram and self-organizing of Kohonen feature map. *Proc. Int. Conf. on Image Processing, KOBE, Japan*, 232-239.
138. Luo, J., Gray, R.T. and Lee, H.C. (1998). Incorporation of derivative priors in adaptive Bayesian color image segmentation. In *Proc. Int. Conf. Image Processing*, 3, 780-784, Chicago.
139. Manjunath, B.S., Symchony, T. and Chellappa, R. (1990). Stochastic and deterministic networks for texture segmentation. *IEEE Trans. Acoustic, Speech, Signal Processing*, 38(6), 1039-1049.
140. Manjunath, B.S. and Ma, W.Y. (1996). Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Analysis and Machine Intell.*, 18(8), 837-842.
141. Martchenko, A.S., Ermolov, I.L., Groumpos, P.P., Poduraev, J.V. and Stylios, C.D. (2003) Investigating Stability analysis Issues for Fuzzy Cognitive Maps, 11th Mediterranean Conference on control and Automation, 1-6 (CD-ROM)
142. Mas, J.F. and Flores, J.J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *Int. J. Remote Sensing*, 29(3), 617-663.
143. Matlab (2009). The Matworks <http://www.mathworks.com/>.
144. MCS (2000-2009). Int. Workshop on Multiple Classifier Systems <http://www.diee.unica.it/mcs/mcs20xx> {xx = 00, 01, 02, 03, 04, 05, 07, 09}.
145. Melas, D.E. and Wilson, S.P. (2002). Double Markov random fields and Bayesian image segmentation. *Signal Processing, IEEE Transactions on*, 50 (2), 357-365.
146. Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sensing*, 42(8), 1178-1190.
147. Miao, Y. and Liu, Z.Q. (2000). On Causal Inference in Fuzzy Cognitive Maps. *IEEE Trans. Fuzzy Systems*, 8(1), 107-119.
148. Mohaddamzadeh A. and Bourbakis, N. (1997). A fuzzy region growing approach for segmentation of color images. *Pattern Recognition*, 30(6), 867-881.
149. Montiel, E., Aguado, A.S. and Nixon, M.S. (2005). Texture classification via conditional histograms. *Pattern Recognition Letters*, 26, 1740-1751.
150. Muñoz-Marí, J., Bruzzone, L. and Camps-Valls, G. (2007). A Support Vector Domain Description Approach to Supervised Classification of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.*, 45(8), 2683-2692.
151. NASA (2009). National Aeronautics and Space Administration, <http://www.nasa.gov/>.

152. Newsam, S., Wang, L. Bhagavathy, S. and Manjunath, B.S. (2004). Using Texture to analyze and Manage Large Collections of Remote Sensed image and Video Data. *Applied Optics*, 43(2), 210-217.
153. Ohio University Brodatz texture image database (2009). <http://sampl.ece.ohio-state.edu/data/stills/texture/Brodatz/index.htm>
154. Özden, M. and Polat, E. (2007). A color image segmentation approach for content-based image retrieval, *Pattern Recognition*, 40(4), 1318-132.
155. Palm, C. and Lehamann, T.M. (2002). Classification of color textures by Gabor filtering. *Machine Graphics and Vision*, 11(2/3), 195-219.
156. Palmer, S.E. (1999). *Vision Science*. Cambridge, MA: MIT Press.
157. Pajares, G., de la Cruz, J. M. and Moreno, V. (2001). Clasificación de texturas mediante redes neuronales, *Ingeniería Civil. Centro de Estudios y Experimentación de Obras Públicas (CEDEX).*- Ministerio de Fomento, 123, 61-69.
158. Pajares, G. and Cruz, J. M. (2002). Clasificación de Texturas Naturales mediante K-Means, *Revista Electrónica de Visión por Computador* <http://revc.uab.es/revista/06/> ISSN 1575-5258), no. 6, 1-18, Ed. Centre de Visió per Computador, Universitat Autònoma Barcelona.
159. Pajares, G., de la Cruz, J. M. and Moreno, V. (2002). Clasificación de texturas naturales mediante agrupamiento borroso. *Ingeniería Civil. Centro de Estudios y Experimentación de Obras Públicas (CEDEX).*- Ministerio de Fomento, 127, 83-89.
160. Pajares, G. (2006). A Hopfield Neural Network for Image Change Detection. *IEEE Trans. Neural Networks*, 17(5), 1250-1264.
161. Pajares, G., Tellaeche, A., BurgosArtizzu, X. and Ribeiro, A. (2007) Design of a computer vision system for a differential spraying operation in precision agriculture through Hebbian learning. *IET Computer Vision*, 1 (3-4), 93-99
162. Pajares, G. and Cruz, J.M. (2007a). *Visión por Computador: Imágenes Digitales y Aplicaciones*, RA-MA, Madrid.
163. Pajares, G. and Cruz, J.M. (2007b). *Ejercicios resueltos de Visión por Computador*, RA-MA, Madrid.
164. Pajares, G. and Santos, M. (2005). *Inteligencia Artificial e Ingeniería del Conocimiento*, RA-MA, Madrid.
165. Pandya, A.S. and Macy, R.B. (1996). *Pattern Recognition with Neural Networks in C++*. CRC Press and IEEE Press, Boca Raton, FL, 410pp.
166. Paola, J.D. and Schowengerdt, R.A. (1995). A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *Int. J. Remote Sensing*, 16(16), 3033-3058.
167. Pappas, T.N. (1992). An adaptive clustering algorithm for image segmentation. *IEEE Trans. Signal Processing*, 40(4), 901-914.
168. Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, 1065-1076.
169. Permuter, H., Francos, J. and Jermyn, I. (2006). A study of Gaussian mixture models of color and texture features for image classification and segmentation. *Pattern Recognition*, 39, 695-706.
170. Petrakos, M., Benediktsson, J. A. and Kanellopoulos, I. (2001). The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion. *IEEE Trans. Geosci. Remote Sens.*, 39(11), 2539–2546.

171. PCI Geomatics (2009). <http://www.pci.on.ca/>.
172. Pi, M.H., Tong, C.S., Choy, S.K. and Zhang, H. (2006). A fast and effective model for wavelet subband histograms and its applications in texture image retrieval. *IEEE Trans. Image Processing*, 15(10), 3078-3088.
173. Porat, M. and Zeevi, Y.Y. (1989). Localized texture processing in vision: analysis and synthesis in the Gaborian space. *IEEE Transactions on Biomedical Engineering (BME)* 36:11, 115-129, 1989
174. Precioso, F.M., Blu Barlaud, T. and Unser, M. (2005). Robust real time segmentation of images and videos using a smooth-spline snake-based algorithm. *IEEE Trans. Image Processing*, 14(7), 910-924.
175. Precioso, F.M., Blu Barlaud, T. and Unser, M. (2005). Robust real time segmentation of images and videos using a smooth-spline snake-based algorithm. *IEEE Trans. Image Processing*, 14(7), 910-924.
176. Proespacio (2009). Agrupación de empresas. <http://www.proespacio.org/>.
177. Puig, D., and García, M.A. (2006). Automatic texture feature selection for image pixel classification. *Pattern Recognition*, 39 (11), 1996-2009.
178. Pyun, K., Lim, J. Won, C.S. and Gray, R.M. (2007). Image Segmentation using Hidden Markov Gauss mixture models. *IEEE Trans. Image Processing*, 16(7), 1902-1911.
179. Pyun, K., Won, C.S., Lim, J. and Gray, R.M. (2002). Texture classification based on multiple Gauss mixture vector quantizer. In *Proc. of ICME*, 501-504.
180. Randen, T. and Husøy, J.H. (1999). Filtering for Texture Classification: A Comparative Study. *IEEE Trans. Pattern Analysis Machine Intell.*, 21(4), 291-310.
181. Ríos-Insúa, S., Bielza, C., y Mateos, A. (2002) Fundamentos de los Sistemas de Ayuda a la Decisión. RA-MA.
182. Robinson, G.S. (1977). Color edge detection. *Optical Engineering*, 16(5), 479-484.
183. Rosenfeld, A., Hummel, R. and Zucker, S. (1976). Scene labeling by relaxation operation. *IEEE Trans. Systems Man Cybernetics*, 6, 420-453.
184. Rud, R., Shoshany, M., Alchanatis, V. and Cohen, Y. (2006). Application of spectral features' ratios for improving classification in partially calibrated hyperspectral imagery: a case study of separating Mediterranean vegetation species. *Journal Real-Time Image Processing*, 1, 143-152.
185. SAR (2009). Synthetic Aperture Radar, Jet Propulsion Laboratory. <http://southport.jpl.nasa.gov/>.
186. Sarkar, A., Biswas, M.K. and Sharma, K.M.S. (2000). A simple unsupervised MRF model based image segmentation approach. *Image Processing, IEEE Transactions on*, 9 (5), 801-812.
187. Seong, J.C. and Usery, E.L. (2001). Fuzzy image classification for continental scale multitemporal NDVI series images using invariant pixels and an image stratification method. *Photogrammetric Engineering and Remote Sensing*, 67(3), 287-294.
188. Sharif, A.M. and Irani, Z. (2005). Exploring Fuzzy Cognitive Mapping for IS Evaluation. *European Journal of Operational Research*, 173, 1175-1187.
189. Shorliffe, E.H. (1976). *Computer-Based Medical Consultations: MYCIN*. Elsevier, New York, NY.
190. SIC, Satellite Image Corporation (2009). <http://www.satimagingcorp.com/>.

191. Singh, S., Haddon, J. and Markou, M. (2001). Nearest-neighbour classifiers in natural scene analysis. *Pattern Recognition*, 34, 1601-1612.
192. Smith, G. and Burns, I. (1997). Measuring texture classification algorithms. *Pattern Recognition Letters*, 18, 1495-1501.
193. Stach, W., Kurgan, L. and Pedrycz, W. (2005). Reformat M. Genetic learning of fuzzy cognitive maps. *Fuzzy Sets and Systems*, 153(3), 371-401.
194. Tax, D.M.J., van Breukelen, M., Duin, R.P.W. and Kittler, J. (2000). Combining multiple classifiers by averaging or by multiplying?. *Pattern Recognition*, 33, 1475-1485.
195. Tellaeche, A., BurgosArtizzu, X.P., Pajares, G. and Ribeiro, A. (2008). A vision-based method for weeds identification through the Bayesian decision theory. *Pattern Recognition*, 41(2), 521-530.
196. Titterton, D.M., Murray, G.D., Murray, L.S., Spiegelhalter, D.J., Skene, A.M., Habbema, J.D.F. and Gelpke, G.J. (1981). Comparison of discriminant techniques applied to a complex data set of head injured patients. *Journal of the Royal Statistical Society, Series A*, 144, 2, 145-175.
197. Tkalcic, J.T.M. (2003). Colour spaces-perceptual, historical and applicational background. In Zajc, M.T.B (Eds.) *Eurocom 2003 Proceedings*, IEEE Region, 8, 1-8.
198. Tottrup, C. (2007). Forest and Land Cover Mapping in a Tropical Highland Region. *Photogrammetric Engineering and Remote Sensing*, 73(9), 1057-1066.
199. Trauwaert, E., Kaufman, L. and Rousseeuw, P. (1991). Fuzzy Clustering Algorithms Based on the Maximum Likelihood Principle. *Fuzzy Sets and Systems*, 42, 213-227.
200. Trias-Sanz, R., Stamon, G. and Louchet, J. (2008). Using colour, texture, and hierarchical segmentation for high-resolution remote sensing. *ISPRS Journal of Photogrammetry & Remote Sensing*, 63, 156-168.
201. Tsardias, A.K. and Margaritis, K.G. (1997). Cognitive Mapping and Certainty Neuron Fuzzy Cognitive Maps. *Information Sciences* 101, 109-130.
202. Tsardias, A.K. and Margaritis, K.G. (1998). The MYCIN certainty factor handling as uniform operator and its use as threshold function in artificial neurons. *Fuzzy Sets and Systems*, 93, 263-274.
203. Tsardias, A.K. and Margaritis, K.G. (1999). An experimental study of the dynamics of the certainty neuron fuzzy cognitive maps. *Neurocomputing*, 24, 95-116.
204. Tso, B.C.K. and Mather, P.M. (1999). Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields. *IEEE Trans. Remote Sensing*, 37(3), 1255-1260.
205. Tubbs, J.D. and Alltop, W.O. (1991). Measures of confidence associated with combining classification rules. *IEEE Trans. Systems, Man, and Cybernetics*, 21, 690-692.
206. Unser, M. (1995). Texture classification and segmentation using wavelet frames. *IEEE Trans. Image Processing*, 4(11), 1549-1560.
207. Valdovinos, R.M., Sánchez, J.S. and Barandela, R. (2005). Dynamic and Static weighing in classifier fusion, in: J.S. Marques, N. Pérez de la Blanca and P. Pina (Eds.), *Pattern Recognition and Image Analysis, Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 59-66.
208. Valdovinos, R.M., and Sánchez, J.S. (2007). Performance Analysis of Classifier Ensembles: Neural Networks Versus Nearest Neighbor Rule. (J. Martí, J.M. Benedí. A.M. Mendonça and J. Serrat (Eds.)). *Pattern Recognition and Image Analysis. Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 4477, Part I, 105-112.

209. Vazquez, E., Baldrich, R. Vazquez, J. and Vanrell, M. (2007). Topological Histogram Reduction towards colour segmentation. (J. Martí, J.M. Benedí, A.M. Mendonça and J. Serrat (Eds.)). Pattern Recognition and Image Analysis. Lecture Notes in Computer Science, Springer-Verlag, Berlin, 4477, Part I, 55-62.
210. Verikas, A. and Lipnickas, A. (2002). Fusing neural networks through space partitioning and fuzzy integration. Neural Processing Letters, 16, 53-65.
211. Verikas, A., Lipnickas, A. Malmqvist, K. Bacauskiene, M. and Gelzinis, A. (1999). Soft combination of neural classifiers: A comparative study. Pattern Recognition Letters, 20, 429-444.
212. VisTex (2009). <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>
213. Volkovich, Z., Barzily, Z. and Morozensky, L. (2008). A statistical model of cluster stability. Pattern Recognition, 41(7), 2174-2188.
214. Wagner T. (1999). Texture Analysis, in: B. Jähne, H. Haußecker, P. Geißler (Eds.), Handbook of Computer Vision and Applications, Academic Press, 2 (Signal Processing and Pattern Recognition).
215. Wangenheim, A., Bertoldi, R.F., Abdala, D.D. and Richter, M.M. (2007). Color image segmentation guided by a color gradient network. Pattern Recognition Letters, 28, 1795-1803.
216. Wang, D. (2005). The time dimension for scene analysis. IEEE Trans. Neural Networks, 16(6), 1401-1426.
217. Wang, Z. and Boesch, R. (2007). Color- and Texture-Based Image Segmentation for Improved Forest Delineation. IEEE Trans. Geoscience and Remote Sensing, 45(10), 3055-3062.
218. Wang, W. and Fenton, N. (2006). Risk and confidence analysis for fuzzy multicriteria decision making. Knowledge Based Systems, 19, 430-437.
219. Wan, T., Canagarajah, N. and Achim, A. (2007). Multiscale Color-Texture image segmentation with adaptive Region Merging. In Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP08), 1, 1-1213-1216.
220. Wilkinson, G. G. (2005). Results and implications of a Study of Fiteen Years of Satellite Image Classification Experiments. IEEE Trans. Geoscience Remote Sensing, 45(3), 433-440.
221. Wilkinson, G.G., Fierens, F. and Kanellopoulos, I. (1995). Integration of neural and statistical approaches in spatial data classification. Geograph. Syst., 32, 1-20.
222. Wozniak, M., Jackowski, K. (2008). Fusers Based on Classifier Response and discriminant Function-Comparative Study. (E. Corchado, A. Abraham and W. Pedrycz (Eds.)). Hybrid Artificial Intelligent Systems. Lecture Notes in Artificial Intelligent, Springer-Verlag, Berlin, 5271, 361-368.
223. Xia, Y., Feng, D., Wang, T., Zhao, R. and Zhang, Y. (2007). Image segmentation by clustering of spatial patterns. Pattern Recognition Letters, 28, 1548-1555.
224. Xia, Y., Feng, D. and Zhao, R. (2006). Adaptive Segmentation of Textured Images by using the coupled Markov Random Field Model. IEEE Trans. Image Processing, 15(11), 3559-3566.
225. Xie, J., and Tsui, H.T. (2004). Image Segmentation based on maximum-likelihood estimation and optimum entropy-distribution (MLE-OED). Pattern Recognition Letters, 25(10), 1133-1141.

- 226. Xu, L., Krzyzak, A. and Suen, C.Y. (1992). Methods of combining multiple classifiers and their application to handwriting recognition. *IEEE Trans. System, Man and Cybernetics*, 22, 418-435.
- 227. Yager, R.R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans. System, Man and Cybernetics*, 18(1), 183-190.
- 228. Yang, A.Y., Wright, J., Ma, Y. and Sastry, S.S. (2008). Unsupervised segmentation of natural images via lossy data compression, *Computer Vision and Image Understanding*, 110, 212-225.
- 229. Yang, F. and Jiang, T. (2003). Pixon-Based Image Segmentation With Markov Random Fields. *IEEE Transactions on image processing*, 12, 12.
- 230. Yu, H., Li, M., Zhang, H.J. and Feng, J. (2002). Color texture moments for content-based image retrieval, in: *Proc. International Conference on Image Processing*, 3, 24-28.
- 231. Yu, R. and Tzeng, G. (2006). A soft computing method for multi-criteria decision making with dependence and feedback. *Applied Mathematics and Computation*, 180(1), 63-75.
- 232. Zadeh L.A. (1975). The concept of a linguistic variable and its application to approximate reasoning. *Inform. Sci.* 8 (1975) 199-249(I), 301-357(II).
- 233. Zhang, J. and Foody, G.M. (1998). A fuzzy classification of sub-urban land cover from remotely sensed imagery. *Int. J. Remote Sensing*, 19(14), 2721-2738.
- 234. Zhang, Y. and Wang, R. (2004). A combined method for texture analysis and its application. F. Bubak, M. (Eds.). *Lecture Notes in Computer Science*, 3036/2004, 413-416.
- 235. Zimmermann, H.J. (1991). *Fuzzy set theory and its applications*, Kluwer Academic Publishers, Norwell, MA.
- 236. Zuyuan, W. and Boesch, R. (2007). Color- and Texture-Based Image Segmentation for Improved Forest Delineation. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10), 3055-3062.